

THIS WEEK

EDITORIALS

WORLD VIEW The lasting value of the doubtful claim about the speed of neutrinos **p.125**

FLU Another way for the virus to take wing **p.127**



TRENDS China helps to drive decline in the world's poorest **p.129**

Lessons of a triple disaster

The aftermath of the biggest earthquake in Japan's history, and the tsunami and nuclear disaster that followed, offers a map for preparing for the next catastrophe.

It is easy to find fault when more than 15,000 people are dead or missing and thousands of square kilometres of countryside are contaminated with radioactivity. Japan's authorities, nuclear industry and scientists share the blame for the disaster that followed the giant earthquake and tsunami that struck the Tohoku coast a year ago, on 11 March 2011. However, many more lives would have been lost had it not been for the work of scientists, engineers, emergency officials and a well-prepared populace. The lessons from what went wrong and right should guide Japan as it rebuilds itself from the strongest earthquake in its history and the nuclear meltdown that followed.

First, what went wrong. The articles and commentaries in this week's special issue (see page 137) show how Japanese scientists and authorities overlooked evidence that the Tohoku coastline was prone to much larger earthquakes and tsunamis than history predicted. Coastal cities and crucial facilities, such as the Fukushima Daiichi nuclear power plant, therefore missed the opportunity to strengthen their defences in advance. At first, the Japan Meteorological Agency (JMA) underestimated the size of the quake and of the tsunami, creating a false sense of security. And coastal residents did not take enough responsibility: one survey of survivors found that 40% who had received a tsunami warning had waited before evacuating.

Japan also failed to make the best use of its technology. Underwater sensors along the Tohoku coastline detected the tsunami when it was 50 kilometres offshore, but precious minutes were lost because mechanisms were not in place to make use of the initial readings. The waves were hitting by the time the JMA corrected its underestimate.

Such criticisms should not overshadow what went right. Relatively few people died during the earthquake — a testament to Japan's rigorous building codes and emergency systems. And the JMA deserves praise for issuing a tsunami warning within three minutes of the earthquake's start. Although it underestimated the size, the warning gave perhaps hundreds of thousands of people a chance to save themselves.

There is plenty of room for improvement, however. Geoscientists must do a better job of deciphering Japan's earthquake history and monitoring current ground deformation to work out which areas are prone to damaging quakes. Researchers and civil authorities are installing new earthquake and tsunami sensors, and they must learn to analyse the data more rapidly. New tsunami simulations, which will be crucial to rebuilding the Tohoku region and fortifying other coastal areas, should be made transparent and be reviewed by external experts. And social scientists can examine how people respond to hazard information — both before and during a crisis — which should help in preparing for future disasters, in Japan and elsewhere.

The meltdowns at the Fukushima Daiichi nuclear plant also revealed both strengths and shortcomings. Despite chaos on the ground, the Japanese government managed to evacuate tens of thousands of residents living within 20 kilometres of the plant in a matter of hours. Health studies indicate that the quick response almost certainly prevented most

people from receiving a significant dose of radiation. Those studies also speak of another strength of the Japanese government — the ability to conduct extensive surveys of affected areas and populations.

Meanwhile, a skilled workforce stayed inside the plant and managed to limit the damage. Much has been made of bureaucratic stumbling in the first hours of the crisis, but despite blackout conditions inside the plant, highly trained staff managed to flood the reactors with water and neutron-absorbing boric acid. Later, scientists and engineers improvised a filtration system to decontaminate cooling water. With repeated leaks and outages, it is far from perfect, but still an impressive solution.

These technical achievements do not offset the government's failure to communicate the dangers of the emergency. Initially, officials downplayed the crisis. Even after the full scale of the meltdowns became clear, radiation measurements were often reported without context, and some were simply wrong. The government also failed to release some information: computer models showing the real-time spread of radioactivity were withheld from the public until 12 days after the crisis began.

The inadequate tsunami warnings cost lives within minutes, but the damage from the failings at Fukushima will stretch on. Sloppy and incomplete reporting has fed public mistrust of the government and its scientists (see page 138), which will hamper efforts both to study and to recover from the crisis. ■

"The inadequate tsunami warnings cost lives within minutes, but the damage from the failings at Fukushima will stretch on."

Political science

The practice of science cannot be, nor should it be, entirely apolitical.

Paul Nurse, president of Britain's Royal Society, does not think he is sitting in an ivory tower, and he has made it clear that he considers that scientists have duties to fulfil and battles to fight beyond the strictly scientific — for example to "expose the bunkum" of politicians who abuse and distort science. This was evident again last week, when Nurse delivered the prestigious Dimbleby Lecture in London, instituted in memory of the British broadcaster Richard Dimbleby.

Nurse identified support for the National Health Service, the need for an immigration policy that attracts foreign scientists, and inspirational science teaching in primary education as some of the priorities

THIS WEEK

EDITORIALS

WORLD VIEW The lasting value of the doubtful claim about the speed of neutrinos **p.125**

FLU Another way for the virus to take wing **p.127**



TRENDS China helps to drive decline in the world's poorest **p.129**

Lessons of a triple disaster

The aftermath of the biggest earthquake in Japan's history, and the tsunami and nuclear disaster that followed, offers a map for preparing for the next catastrophe.

It is easy to find fault when more than 15,000 people are dead or missing and thousands of square kilometres of countryside are contaminated with radioactivity. Japan's authorities, nuclear industry and scientists share the blame for the disaster that followed the giant earthquake and tsunami that struck the Tohoku coast a year ago, on 11 March 2011. However, many more lives would have been lost had it not been for the work of scientists, engineers, emergency officials and a well-prepared populace. The lessons from what went wrong and right should guide Japan as it rebuilds itself from the strongest earthquake in its history and the nuclear meltdown that followed.

First, what went wrong. The articles and commentaries in this week's special issue (see page 137) show how Japanese scientists and authorities overlooked evidence that the Tohoku coastline was prone to much larger earthquakes and tsunamis than history predicted. Coastal cities and crucial facilities, such as the Fukushima Daiichi nuclear power plant, therefore missed the opportunity to strengthen their defences in advance. At first, the Japan Meteorological Agency (JMA) underestimated the size of the quake and of the tsunami, creating a false sense of security. And coastal residents did not take enough responsibility: one survey of survivors found that 40% who had received a tsunami warning had waited before evacuating.

Japan also failed to make the best use of its technology. Underwater sensors along the Tohoku coastline detected the tsunami when it was 50 kilometres offshore, but precious minutes were lost because mechanisms were not in place to make use of the initial readings. The waves were hitting by the time the JMA corrected its underestimate.

Such criticisms should not overshadow what went right. Relatively few people died during the earthquake — a testament to Japan's rigorous building codes and emergency systems. And the JMA deserves praise for issuing a tsunami warning within three minutes of the earthquake's start. Although it underestimated the size, the warning gave perhaps hundreds of thousands of people a chance to save themselves.

There is plenty of room for improvement, however. Geoscientists must do a better job of deciphering Japan's earthquake history and monitoring current ground deformation to work out which areas are prone to damaging quakes. Researchers and civil authorities are installing new earthquake and tsunami sensors, and they must learn to analyse the data more rapidly. New tsunami simulations, which will be crucial to rebuilding the Tohoku region and fortifying other coastal areas, should be made transparent and be reviewed by external experts. And social scientists can examine how people respond to hazard information — both before and during a crisis — which should help in preparing for future disasters, in Japan and elsewhere.

The meltdowns at the Fukushima Daiichi nuclear plant also revealed both strengths and shortcomings. Despite chaos on the ground, the Japanese government managed to evacuate tens of thousands of residents living within 20 kilometres of the plant in a matter of hours. Health studies indicate that the quick response almost certainly prevented most

people from receiving a significant dose of radiation. Those studies also speak of another strength of the Japanese government — the ability to conduct extensive surveys of affected areas and populations.

Meanwhile, a skilled workforce stayed inside the plant and managed to limit the damage. Much has been made of bureaucratic stumbling in the first hours of the crisis, but despite blackout conditions inside the plant, highly trained staff managed to flood the reactors with water and neutron-absorbing boric acid. Later, scientists and engineers improvised a filtration system to decontaminate cooling water. With repeated leaks and outages, it is far from perfect, but still an impressive solution.

These technical achievements do not offset the government's failure to communicate the dangers of the emergency. Initially, officials downplayed the crisis. Even after the full scale of the meltdowns became clear, radiation measurements were often reported without context, and some were simply wrong. The government also failed to release some information: computer models showing the real-time spread of radioactivity were withheld from the public until 12 days after the crisis began.

The inadequate tsunami warnings cost lives within minutes, but the damage from the failings at Fukushima will stretch on. Sloppy and incomplete reporting has fed public mistrust of the government and its scientists (see page 138), which will hamper efforts both to study and to recover from the crisis. ■

"The inadequate tsunami warnings cost lives within minutes, but the damage from the failings at Fukushima will stretch on."

Political science

The practice of science cannot be, nor should it be, entirely apolitical.

Paul Nurse, president of Britain's Royal Society, does not think he is sitting in an ivory tower, and he has made it clear that he considers that scientists have duties to fulfil and battles to fight beyond the strictly scientific — for example to "expose the bunkum" of politicians who abuse and distort science. This was evident again last week, when Nurse delivered the prestigious Dimbleby Lecture in London, instituted in memory of the British broadcaster Richard Dimbleby.

Nurse identified support for the National Health Service, the need for an immigration policy that attracts foreign scientists, and inspirational science teaching in primary education as some of the priorities

for British scientists. These, and many other issues he raised, such as increasing scientists' interactions with industry, commerce and the media, and resisting the politicization of climate-change research, are relevant throughout the world and not just in Britain.

All the more reason not to misinterpret Nurse's insistence on a separation of science and politics: as he put it, first we need the science, then the politics. What Nurse rightly warned against is the intrusion of ideology into the interpretation and acceptance of scientific knowledge as, for example, in the Soviet Union's support of the anti-Mendelian biology of Trofim Lysenko. Given recent accounts of political interference in climate research in the United States (N. Oreskes and E. M. Conway *Nature* 465, 686; 2010), this is a timely reminder.

But it is easy to render this equation too simplistically. For example, Nurse also cited the rejection by Adolf Hitler of Albert Einstein's relativistic physics as 'Jewish physics'. But that is not quite how it was. 'Jewish physics' was a straw man invented by the anti-Semitic and pro-Nazi physicists Johannes Stark and Philipp Lenard, partly because of professional jealousies and grudges. The Nazi leaders were, however, largely indifferent to what looked like an academic squabble, and in the end lost interest in Stark and Lenard's risible 'Aryan physics' because they needed a physics that actually worked.

That is one reason to be sceptical of the common claim, repeated by Nurse, that science can flourish only in a free society. Historians of science in Nazi Germany such as Kristie Macrakis (in her book *Surviving the Swastika*) have challenged this assertion, which is not made true simply because we would like it to be so. Authoritarian regimes are perfectly capable of putting pragmatism before ideology. The scientific process itself is not impeded by state control in China — quite the contrary — and the old canard that Chinese science lacks innovation and daring is now transparently nonsense. During the cold war, some Soviet science was vibrant and bold. Even the most notorious example of state repression of science — the trial of Galileo — is

apt to be portrayed too simplistically as a conflict of faith and reason rather than a collision of personalities and circumstances (none of which exonerates Galileo's scandalous persecution).

There is a more compelling lesson to be drawn from Nazi Germany that bears on Nurse's themes: although political (and religious) ideology has no place in deciding scientific questions, the practice of science is inherently political. In that sense, science can never come before politics. Scientists everywhere enter into a social contract, not least because they are not their own paymasters. Much, if

not most, scientific research has social and political implications, often broadly visible from the outset. In times of crisis (like the present), scientists must respond intellectually and professionally to the challenges facing society, and not think that safeguarding their funding is enough.

The consequences of imagining that science can remain aloof from politics became acutely apparent in Germany in 1933, when the consensus view that politics was, as Heisenberg put it, an unseemly "money business" meant that most scientists saw no reason to mount concerted resistance to the expulsion of Jewish colleagues — regarded as a political rather than a moral matter. This 'apolitical' attitude can now be seen as a convenient myth that led to acquiescence in the Nazi regime and made it easy for German scientists to be manipulated. It would be naive to imagine that only totalitarianism could create such a situation.

The rare and most prominent exception to apolitical behaviour was Einstein, whose outspokenness dismayed even his principled friends the German physicists Max Planck and Max von Laue. "I do not share your view that the scientist should observe silence in political matters," he told them. "Does not such restraint signify a lack of responsibility?" There was no hint of such a lack in Nurse's talk. But we must take care to distinguish the political immunity of scientific reasoning from the political dimensions and obligations of doing science. ■

Gold in the text?

Publishers and scientists should do more to foster the mining of research literature by computer.

Whether from the petabytes of data produced by the Large Hadron Collider, or the hundreds of millions of bases in the human genome, much scientific analysis nowadays relies on computers to pull out meaning from swathes of data. But one vast store of information, the research literature, has so far seemed immune to computer analysis. By and large, articles exist only in formats designed for humans to read — such as this paragraph.

Text-mining aims to break down this barrier. Using natural-language-processing concepts honed over the past 30 years, computer programs are starting to pull out information from plain text, including patents and research articles. Right now, the software requires highly skilled operators, but in the next decade it might transform the way scientists read the literature. Text-miners hope to make scientific discoveries by scouring hundreds of research papers for associations and connections (such as between drugs and side effects, or genes and disease pathways) that humans reading each paper individually might not notice.

The promise is yet to be backed up with concrete examples of scientific success — although in the pharmaceutical industry, text-mining companies are already working with researchers to speed up drug discovery. But academics are struggling to even run experiments — because publishing licences do not let them text-mine research papers, and publishers are slow to respond to text-mining requests. Fed up after two years of negotiations, one team of researchers is launching a public website to log publishers' responses (see page 134).

There is no doubt that a completely open research literature would make it easier to demonstrate how such machine-reading can lead to scientific discovery. But the question is how to make progress today, when much research lies behind subscription firewalls and even 'open' content does not always come with a text-mining licence (including 83% of the 'free' research in the PubMedCentral online archive).

Publishers should agree that scientists who have already paid for access to research papers may text-mine content at no extra cost and publish their findings — as long as their doing so does not breach the original firewall. Publishers can have no claim on the data in articles, only on the way in which the articles have been edited and formatted. They should make their text-mining policies clear and consider following the example of the journal *Heredity*, which says it is "seeking to encourage text-mining experiments". (Its publisher, Nature Publishing Group, which also publishes this journal, says that it does not charge subscribers to mine content, subject to contract.)

On the other hand, text-miners need to make a better case for their technology. They say they are in a catch-22 situation — how can they demonstrate the benefits if they aren't allowed to run experiments on the literature? Instead, they text-mine abstracts, usually by picking out key words — a pale shadow of what full-text-mining might offer. Casey Bergman at the University of Manchester, UK, is chronicling projects that have tried to text-mine the available PubMedCentral content (see go.nature.com/2pqp8g) and finds very few examples — suggesting that text-miners are reluctant even to mine the corpus of free content.

Publishers point out that they receive few text-mining requests, so the field can't be very hot. So unless text-miners start to make full use of the content that is available, and request more access to published content — while always being clear about how their project will benefit science — the unsatisfactory impasse will continue. ■

NATURE.COM
To comment online,
click on Editorials at:
go.nature.com/xhunjv

for British scientists. These, and many other issues he raised, such as increasing scientists' interactions with industry, commerce and the media, and resisting the politicization of climate-change research, are relevant throughout the world and not just in Britain.

All the more reason not to misinterpret Nurse's insistence on a separation of science and politics: as he put it, first we need the science, then the politics. What Nurse rightly warned against is the intrusion of ideology into the interpretation and acceptance of scientific knowledge as, for example, in the Soviet Union's support of the anti-Mendelian biology of Trofim Lysenko. Given recent accounts of political interference in climate research in the United States (N. Oreskes and E. M. Conway *Nature* 465, 686; 2010), this is a timely reminder.

But it is easy to render this equation too simplistically. For example, Nurse also cited the rejection by Adolf Hitler of Albert Einstein's relativistic physics as 'Jewish physics'. But that is not quite how it was. 'Jewish physics' was a straw man invented by the anti-Semitic and pro-Nazi physicists Johannes Stark and Philipp Lenard, partly because of professional jealousies and grudges. The Nazi leaders were, however, largely indifferent to what looked like an academic squabble, and in the end lost interest in Stark and Lenard's risible 'Aryan physics' because they needed a physics that actually worked.

That is one reason to be sceptical of the common claim, repeated by Nurse, that science can flourish only in a free society. Historians of science in Nazi Germany such as Kristie Macrakis (in her book *Surviving the Swastika*) have challenged this assertion, which is not made true simply because we would like it to be so. Authoritarian regimes are perfectly capable of putting pragmatism before ideology. The scientific process itself is not impeded by state control in China — quite the contrary — and the old canard that Chinese science lacks innovation and daring is now transparently nonsense. During the cold war, some Soviet science was vibrant and bold. Even the most notorious example of state repression of science — the trial of Galileo — is

apt to be portrayed too simplistically as a conflict of faith and reason rather than a collision of personalities and circumstances (none of which exonerates Galileo's scandalous persecution).

There is a more compelling lesson to be drawn from Nazi Germany that bears on Nurse's themes: although political (and religious) ideology has no place in deciding scientific questions, the practice of science is inherently political. In that sense, science can never come before politics. Scientists everywhere enter into a social contract, not least because they are not their own paymasters. Much, if

not most, scientific research has social and political implications, often broadly visible from the outset. In times of crisis (like the present), scientists must respond intellectually and professionally to the challenges facing society, and not think that safeguarding their funding is enough.

The consequences of imagining that science can remain aloof from politics became acutely apparent in Germany in 1933, when the consensus view that politics was, as Heisenberg put it, an unseemly "money business" meant that most scientists saw no reason to mount concerted resistance to the expulsion of Jewish colleagues — regarded as a political rather than a moral matter. This 'apolitical' attitude can now be seen as a convenient myth that led to acquiescence in the Nazi regime and made it easy for German scientists to be manipulated. It would be naive to imagine that only totalitarianism could create such a situation.

The rare and most prominent exception to apolitical behaviour was Einstein, whose outspokenness dismayed even his principled friends the German physicists Max Planck and Max von Laue. "I do not share your view that the scientist should observe silence in political matters," he told them. "Does not such restraint signify a lack of responsibility?" There was no hint of such a lack in Nurse's talk. But we must take care to distinguish the political immunity of scientific reasoning from the political dimensions and obligations of doing science. ■

Gold in the text?

Publishers and scientists should do more to foster the mining of research literature by computer.

Whether from the petabytes of data produced by the Large Hadron Collider, or the hundreds of millions of bases in the human genome, much scientific analysis nowadays relies on computers to pull out meaning from swathes of data. But one vast store of information, the research literature, has so far seemed immune to computer analysis. By and large, articles exist only in formats designed for humans to read — such as this paragraph.

Text-mining aims to break down this barrier. Using natural-language-processing concepts honed over the past 30 years, computer programs are starting to pull out information from plain text, including patents and research articles. Right now, the software requires highly skilled operators, but in the next decade it might transform the way scientists read the literature. Text-miners hope to make scientific discoveries by scouring hundreds of research papers for associations and connections (such as between drugs and side effects, or genes and disease pathways) that humans reading each paper individually might not notice.

The promise is yet to be backed up with concrete examples of scientific success — although in the pharmaceutical industry, text-mining companies are already working with researchers to speed up drug discovery. But academics are struggling to even run experiments — because publishing licences do not let them text-mine research papers, and publishers are slow to respond to text-mining requests. Fed up after two years of negotiations, one team of researchers is launching a public website to log publishers' responses (see page 134).

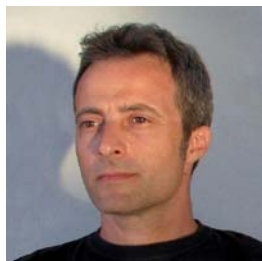
There is no doubt that a completely open research literature would make it easier to demonstrate how such machine-reading can lead to scientific discovery. But the question is how to make progress today, when much research lies behind subscription firewalls and even 'open' content does not always come with a text-mining licence (including 83% of the 'free' research in the PubMedCentral online archive).

Publishers should agree that scientists who have already paid for access to research papers may text-mine content at no extra cost and publish their findings — as long as their doing so does not breach the original firewall. Publishers can have no claim on the data in articles, only on the way in which the articles have been edited and formatted. They should make their text-mining policies clear and consider following the example of the journal *Heredity*, which says it is "seeking to encourage text-mining experiments". (Its publisher, Nature Publishing Group, which also publishes this journal, says that it does not charge subscribers to mine content, subject to contract.)

On the other hand, text-miners need to make a better case for their technology. They say they are in a catch-22 situation — how can they demonstrate the benefits if they aren't allowed to run experiments on the literature? Instead, they text-mine abstracts, usually by picking out key words — a pale shadow of what full-text-mining might offer. Casey Bergman at the University of Manchester, UK, is chronicling projects that have tried to text-mine the available PubMedCentral content (see go.nature.com/2pqp8g) and finds very few examples — suggesting that text-miners are reluctant even to mine the corpus of free content.

Publishers point out that they receive few text-mining requests, so the field can't be very hot. So unless text-miners start to make full use of the content that is available, and request more access to published content — while always being clear about how their project will benefit science — the unsatisfactory impasse will continue. ■

NATURE.COM
To comment online,
click on Editorials at:
go.nature.com/xhunjv



No theory is too special to question

The flurry of research that followed the claim of faster-than-light neutrinos was far from a waste of time, says **Giovanni Amelino-Camelia**.

So, it seems that neutrinos cannot travel faster than light. Since the OPERA collaboration reported six months ago that they could, some sizeable systematic errors have been uncovered, and the result might not stand up to further scrutiny.

The situation has prompted the fundamental-physics community to discuss the proper way to handle cases in which preliminary experimental results challenge 'established' laws. (In this case, one that many physicists hold dearer than most — Einstein's special theory of relativity.)

Some colleagues feel that such results should not be highly publicized, at least until the findings have been properly checked. In hindsight, the frantic attempts to analyse, understand and reproduce the OPERA result may seem unnecessary. But I believe that fundamental physics learned much in those months of madness, as scientists chased the dream of a fundamental revolution.

The unexpected result has inspired more than 200 studies since last September — a huge number, given the usual pace of publication in fundamental physics. In my field of quantum-gravity research — a pretty active field — a result would be considered to have taken the community by storm if it inspired this number of studies over four or five years.

Of course, not all the papers were impressive. But some research programmes produced valuable results in a very short time.

Thanks to Internet communication, the OPERA experiment was scrutinized remarkably quickly, in an unprecedented example of what the twenty-first-century scientific method may look like. Within 48 hours of the announcement, analyses showed that the recorded speed of the OPERA neutrinos was at odds not only with special relativity, but also with other neutrino speed measurements.

I cannot be the only theoretical physicist to be astonished at how the experimentalists were able to arrange within weeks a second run of the experiment, which used shorter neutrino pulses. The results were the same — showing that the statistical methods developed by OPERA to handle the original results were reliable, which could prove valuable for future experiments.

Some of the results on the theory side are relevant to my own research, on the interplay between quantum mechanics and relativity at particle energies much higher than the ones we can probe with accelerators — in particular, at the 'Planckian' energies of the Big Bang. In this ethereal research area, one live debate concerns whether the theories proposed to replace special relativity should preserve the relativistic description

of different observers. Remarkably, the same issue turned out to be important in the investigations of the OPERA result: the implications of superluminal neutrinos for the properties of other particles depend on whether or not a relativistic description of observers is preserved.

As we return to our Planckian ivory towers, we theorists now know that the topics we discuss could one day actually matter. And we get something in return, because some of the results for revising relativity at energies already achieved in accelerators will contribute to our study of the fate of relativity at Planckian energies.

Perhaps most importantly, the OPERA affair pushed researchers from different research areas, some of whom never talk to each other, to share their expertise. For example, very few particle physicists have the chance to appreciate the subtleties of the description of time measurements in quantum mechanics. Yet here, some particle physicists were talking to experts on the conceptual foundations of quantum

mechanics, who are in a better position to assess the nature of the neutrino travel time determined by OPERA. As we get better at determining the travel times of particles in accelerators — and OPERA is a good example of this trend — we will one day need to deal with the slippery aspects of the description of time in quantum mechanics. When we do, we will surely benefit from the shared expertise produced by studies of the OPERA anomaly.

But there is a possible negative outcome, and one that we must prevent.

As in most modern particle-physics experiments, the analyses were 'blinded' — the criteria used, including the estimates of systematic errors, were fine-tuned before the data were looked at. Once the data were analysed, the results were announced without much delay and with no further tweaks. If the preliminary OPERA result

is indeed not confirmed, some are bound to propose that we soften the blind-analysis standards, to guard against the perceived wastefulness of questioning a sound law. Experimentalists who find that their results contrast with 'known' physics might be encouraged to postpone announcement of the results, and first look determinedly, in a non-blinded way, for systematic errors that might make the contrast go away. This would introduce a potentially disastrous bias against important discoveries. Questioning our laws, even on the basis of preliminary experiments, is a healthy exercise. We must assume that the next fundamental physics revolution is just beyond our nose, safely outside the reach of our brains, but within the grasp of the next truly innovative experiment. ■

Giovanni Amelino-Camelia is a theoretical physicist at the University of Rome La Sapienza.
e-mail: giovanni.amelino-camelia@roma1.infn.it

WE WILL SURELY
BENEFIT FROM
THE SHARED
EXPERTISE
PRODUCED BY
STUDIES OF THE
OPERA
ANOMALY.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/ugdsbr

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

HIV

A race to kill or be killed

HIV's prime cellular victim can itself kill affected cells during the early days of infection. And it seems that the more vigorous this response by CD4 T cells is, the greater an HIV-positive person's chance is of being able to maintain a relatively low viral level and the better their disease outcome is likely to be.

Hendrik Streeck at the Ragon Institute in Charlestown, Massachusetts, and his colleagues studied two groups of people with untreated HIV during the first year of infection. By the end of the period, members of one group had developed a higher viral load than had those of the other. The authors found that patients with a lower HIV load had a greater percentage of HIV-specific CD4 T cells. Moreover, at the beginning of the study, a higher proportion of CD4 T cells in these patients expressed proteins, particularly one called granzyme A, that kill other cells.

The authors say that measuring the responses of these T cells during the early stages of infection could allow medics to predict how HIV infection will progress in individual patients.

Sci. Transl. Med. **4**, 123ra25 (2012)

SUNCOR ENERGY



KYODO/REUTERS

OCEANOGRAPHY

Tsunamis collide and grow taller

Ridges and mountains on the sea floor dangerously amplified the eastward-bound segments of the tsunami that devastated Japan's Tohoku coast (pictured) last year.

Satellite observations made on 11 March at different locations over the Pacific Ocean suggest that tsunami height increased as the wavefront raced towards the west coast of America. Tony Song at NASA's Jet Propulsion Laboratory in Pasadena, California, and his team compared satellite data with a tsunami simulation and maps of ocean-bottom topography. They conclude that

topographic barriers such as the Hawaiian Ridge slowed and curved the eastward-bound tsunami front. As segments of this disrupted front collided and merged, the peak-to-trough height of the tsunami wave more than doubled.

Understanding how tsunamis interact with ocean-floor topography could help researchers to improve tsunami forecasting.

Geophys. Res. Lett. <http://dx.doi.org/10.1029/2011GL050767> (2012)

For more on the Japan earthquake, see www.nature.com/japanquake.

ENVIRONMENTAL SCIENCE

Oil-sands pollution quantified

Air pollution caused by crude-oil extraction from Canadian oil sands is comparable to that measured over mid-sized cities or in the vicinity of large coal-burning power plants.

Chris McLinden of Environment Canada in Toronto and his colleagues used satellite observations to determine how the mining and processing of oil sands (pictured) affect air quality in Alberta — home to the world's second-largest crude-oil reserve. The authors measured

greatly increased atmospheric levels of two major pollutants, nitrogen dioxide and sulphur dioxide, over a 30-by-50-kilometre area of intense surface mining. They also noted that nitrogen dioxide pollution in the area had risen steadily since 2005.

Geophys. Res. Lett. <http://dx.doi.org/10.1029/2011GL050273> (2012)

IMMUNOLOGY

Fighting viruses antibody-free

Neutralizing antibodies were thought to be essential to helping the body fight off

viruses, but it turns out that this is not always the case.

Antibodies are made by immune cells called B cells. A team led by Matteo Iannacone and Ulrich von Andrian at Harvard Medical School in Boston, Massachusetts, studied mice engineered to make their B cells incapable of producing antibodies but still able to make a chemical called lymphotoxin $\alpha 1\beta 2$. When the animals were infected with vesicular stomatitis virus, this lymphotoxin caused another type of immune cell, the macrophage, to be preferentially infected by the virus. Infected macrophages secreted another molecule,

type I interferon, which prevented fatal viral infection of the nervous system.

The findings suggest that innate immunity — the immune system's first line of defence, which includes macrophages — has an unappreciated role in combating viral infections. *Immunity* <http://dx.doi.org/10.1016/j.immuni.2012.01.013> (2012)

MATERIALS SCIENCE

Competition looms for graphene

Graphene — atomically thick sheets of carbon atoms arranged in a honeycomb pattern — has received much attention for its exceptional electrical properties, thought to arise from its hexagonal symmetry. But Daniel Malko and his colleagues at the University of Erlangen-Nuremberg in Germany propose that a similar material, graphyne, might be just as special — or even more so.

Graphynes are similar to graphene, the main difference being that they have triple bonds between some of their carbon atoms. Only small snippets of graphyne have been synthesized so far, but the authors performed a simulation showing that, like graphene, various forms of graphyne should have high electrical conductivity. The researchers note that not all graphynes have graphene's hexagonal honeycomb configuration. This suggests that many more two-dimensional materials — with various symmetries and non-carbon structures — could have graphene-like properties. *Phys. Rev. Lett.* 108, 086804 (2012)

CELL BIOLOGY

Sperm steer with calcium

How winding a sperm's path is depends on the degree to which the concentration of calcium ions in the sperm changes.

Eggs attract sperm by

sending out chemical cues that activate calcium-ion release in the sperm's tail, or flagellum. When Luis Alvarez and Benjamin Kaupp at the Center of Advanced European Studies and Research in Bonn, Germany, and their co-workers stimulated calcium surges in the flagella of sea-urchin sperm, the sperm switched from swimming in circles to bouts of near-straight runs interlaced with turns. The curvature of the sperm's movement correlated with how quickly the concentration of calcium changed, regardless of the total concentration.

The researchers propose that two different calcium-ion-binding reactions could allow sperm to sense the rate of change in calcium levels. *J. Cell Biol.* <http://dx.doi.org/10.1083/jcb.201106096> (2012)

NEUROSCIENCE

Behind marijuana memory lapse

Marijuana hampers short-term memory by activating a signalling pathway between neurons and non-neuronal cells called astrocytes, work in mice suggests. This is surprising because astrocytes have mostly been considered to be just support cells.

Giovanni Marsicano at the French National Institute of Health and Medical Research in Bordeaux, Xia Zhang of the University of Ottawa in Canada and their colleagues found that marijuana's psychoactive ingredient, tetrahydrocannabinol (THC), weakens the connections, or synapses, between neurons in the hippocampus — a brain region crucial for memory formation. They tested the effects of THC on mice engineered so that either their neurons or their astrocytes lacked the CB₁ receptor that THC acts on. In mice missing the astrocyte receptors, THC had no effect on hippocampal synapses or working memory, as shown in a behavioural test.

The authors propose that the activation of astrocyte

COMMUNITY CHOICE

The most viewed papers in science

MICROBIOLOGY

Unintended antimicrobial effects

HIGHLY READ
on rsbl.royalsocietypublishing.org
in January

As antibiotic resistance in microorganisms has risen, some scientists have suggested that short, synthetic amino-acid chains, or peptides, that kill invading pathogens could make a new therapeutic weapon. But

Michelle Habets and Michael Brockhurst at the University of Liverpool, UK, highlight serious potential risks of this strategy.

The researchers exposed the pathogen *Staphylococcus aureus* — which causes a range of dangerous infections and has developed resistance to many antibiotics — to increasing levels of a synthetic peptide called pexiganan, a candidate treatment for diabetic leg-ulcer infection. Over several tens of generations, the bacteria evolved resistance to the peptide. Furthermore, some bacteria also evolved resistance to HNP-1, a natural human peptide that is an important part of innate immunity, which represents the immune system's first line of defence.

The authors say that such cross-resistance could undermine the innate immune system's ability to prevent superficial infections from progressing to life-threatening disease.

Biol. Lett. <http://dx.doi.org/10.1098/rsbl.2011.1203> (2012)

CB₁ receptors by THC causes receptors for a compound called AMPA to be removed from the membranes of neurons, which is known to weaken synapses.

Cell 148, 1039–1050 (2012)

VIROLOGY

Bats can carry flu too

Fruit bats in Guatemala harbour a strain of influenza virus, researchers report — expanding the known mammalian reservoir for the largely bird-borne influenza A.

Suxiang Tong of the US Centers for Disease Control and Prevention in Atlanta, Georgia, and her colleagues screened 316 bats from 21 species for a variety of viruses. Viral genome sequencing revealed that three little yellow-shouldered bats (*Sturnira*

lilium; pictured), were positive for influenza A. Analysis of two key proteins that are shared by all influenza strains, as well as other viral genes, showed that this strain — designated H17 — diverged from other known influenza viruses long ago.

The authors say that H17 seems to pose no immediate threat to humans. However,

it is similar enough to other subtypes that genetic exchange with them could produce a virus infectious to other animal species, including humans.

Proc. Natl Acad.

Sci. USA <http://dx.doi.org/10.1073/pnas.1116200109> (2012)

For a longer story on this research, see go.nature.com/fnjy2l

➔ **NATURE.COM**

For the latest research published by Nature visit: www.nature.com/latestresearch



ILAN ELIZKOWICZ

SEVEN DAYS

The news in brief

POLICY

Avian-flu review

In the light of fresh data about two unpublished studies that created ferret-transmissible versions of the avian H5N1 flu virus, the US National Institutes of Health (NIH) has told the National Science Advisory Board for Biosecurity to reconsider the fate of the manuscripts. The board had asked that the research be censored, but flu researchers wanted the work published in full. At a public meeting on 29 February, Anthony Fauci, who heads the NIH's National Institute of Allergy and Infectious Diseases, said he wanted the panel to think again. It has become clear, for example, that ferrets did not die when infected through aerosol transmission. See go.nature.com/wj4ce9 for more.

North Korea talks

North Korea has agreed to suspend nuclear testing, uranium enrichment and development of long-range ballistic missiles in exchange for food aid from the United States. The agreement, announced on 29 February, also allows inspectors from the International Atomic Energy Agency to enter the country's Yongbyon nuclear complex for the first time in more than three years. The diplomatic breakthrough is the first major international political deal made by North Korea's new leader, Kim Jong-un. See go.nature.com/3epxya for more.

China budget boost

China's central government plans to increase spending on science and technology by 12.4% compared with last year's budget proposal, to 229 billion renminbi (US\$36.2 billion), according to a draft budget announced on 5 March at the

annual meeting of China's top legislature, the National People's Congress. That rise was slightly smaller than the overall budget increase (14.1%), but the country's main basic-science funding agency, the National Natural Science Foundation, saw a 26% boost to 32.5 billion renminbi, and the budget earmarked for agricultural science and technology jumped by 53%, to 10.1 billion renminbi.

Primate-hub woes

The death of a cotton tamarin monkey at Harvard's New England Primate Research Center — the facility's third monkey death in six months — has led to the suspension of all new experiments and the resignation of interim

director Fred Wang. Wang said on 2 March that he was stepping down "for personal and professional reasons". The centre, which houses roughly 2,000 monkeys, is based in Scarborough, Massachusetts, and receives about US\$25 million a year from the National Institutes of Health. It conducts research on conditions such as AIDS, Parkinson's disease and cancer.

Genetic tests online

The US National Institutes of Health (NIH) has unveiled its Genetic Testing Registry, an online resource for information on genetic tests. The NIH says that genetic tests are now available for some 2,500 diseases, and some can be purchased directly from the

spill to some \$30 billion; it made \$23.9 billion in profits last year. BP does not think that, in total, it will spend more than the \$37.2 billion already set aside for oil-spill costs — but that depends on the extent of pollution claims (and possible criminal charges) from the US government.



B. BESAW/AURORA PHOTOS/CORBIS

The cost of the Gulf of Mexico oil spill

Oil company BP has agreed to pay about US\$7.8 billion to settle claims for damages from an estimated 116,000 businesses and individuals affected by the 2010 oil spill in the Gulf of Mexico (pictured). The deal, outlined on 3 March, brings the firm's total spend on the oil

providers. The information in the registry, launched on 29 February, is voluntarily provided by test producers, and is not independently verified by the NIH. See go.nature.com/zrk56v for more.

RESEARCH

Diesel cancer link

After almost two decades of legal and political wrangling, US government researchers have published research showing how exposure to diesel exhaust increases the risk of lung cancer. Although it is not a surprising finding, the prolonged legal battle over the study's data, which were contested by an industry coalition, attracted wide

attention. An appeals-court ruling paved the way for final publication on 5 March. In June, international regulatory agencies will meet to reassess the carcinogenic effects of diesel exhaust. See go.nature.com/hickkz for more.

Exoplanet update

NASA's Kepler team on 2 March publicized the data for 1,091 signals that may represent extrasolar planets — bringing the total public catalogue to 2,321 candidates discovered during the mission's first 16 months of operation. With the rate of false positives estimated to be less than 10%, these candidates will bolster the database of 760 exoplanets confirmed so far. Some 246 of the Kepler candidates are Earth-sized.

Safer water supplies

More than 2 billion people gained access to improved sources of drinking water (such as piped water or protected wells) between 1990 and 2010, says a report released by the World Health Organization and the United Nations Children's Fund (UNICEF) on 6 March. The report says that half of a Millennium Development Goal has been reached (the full goal is to halve from 1990 levels the proportion of people without sustainable access to safe drinking water and basic sanitation). But

many 'improved' water sources might not actually be safe or sustainable, the report acknowledged.

Ghostwriting row

Two psychiatrists accused last July of putting their names to a 2001 paper that they did not write have been cleared by their institution, the University of Pennsylvania in Philadelphia. The paper was about the antidepressant Paxil (paroxetine), made by the drug giant GlaxoSmithKline. The university says that there is no merit to the misconduct allegations and adds that although the authors had failed to acknowledge the assistance of a medical writing company, current rules that forbid this practice were not in place then. The US Office of Research Integrity, which oversees US Public Health Service agencies and grant recipients, says that it is still considering the matter. See go.nature.com/j2gglf for more.

FUNDING

US research ships

The US Navy says that it has finalized contracts for two new ocean-class research vessels. The ships, costing a total of around US\$145 million, are set to launch in 2014 and 2015. One will be operated by the Scripps Institution



of Oceanography in La Jolla, California, and the other by the Woods Hole Oceanographic Institution in Massachusetts. They will replace Scripps' RV *Melville* and Woods Hole's RV *Knorr* (pictured) — both larger vessels that were launched in 1968 and will be decommissioned. See go.nature.com/pdslvy for more.

PEOPLE

Bioethics conflict

A bioethicist who remained editor-in-chief of a high-impact bioethics journal while working for a controversial stem-cell firm resigned his industry position on 28 February, after criticism from his peers about conflicts of interest. Glenn McGee, who founded the *American Journal of Bioethics*, had been working with Celltex Therapeutics in Houston, Texas, since December 2011 as president for ethics and strategic initiatives. McGee

COMING UP

11 MARCH

Japan commemorates the first anniversary of the earthquake and tsunami that struck the country in 2011. See page 137 for more.

12 MARCH

At the World Water Forum in Marseille, France, the United Nations launches the fourth edition of its World Water Development Report, a triennial study into managing global freshwater resources. worldwaterforum6.org

resigned his editorship in February, but Celltex has been mired in controversy owing to evidence that it is involved in the clinical use of unproven stem-cell treatments in the United States. See go.nature.com/angubk for more.

Mann hunt over

Virginia's supreme court threw out an investigation into climate scientist Michael Mann on 2 March. The state's attorney general, Ken Cuccinelli, had spent almost two years demanding that the University of Virginia in Charlottesville, where Mann once worked, provide him with documentation including the researcher's e-mail records, following the controversy that surrounded climate e-mails leaked from the University of East Anglia in Norwich, UK. Mann, now at Pennsylvania State University in University Park, still faces a separate lawsuit over access to his e-mails, filed by the conservative American Tradition Institute, based in Washington DC. See go.nature.com/xj5czz for more.

► **NATURE.COM**

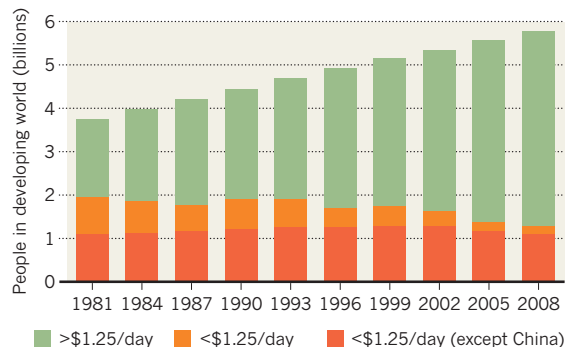
For daily news updates see:
www.nature.com/news

TREND WATCH

The developing world has met the United Nations' first Millennium Development Goal, to halve extreme poverty from 1990 levels by 2015, the World Bank said on 29 February on the basis of preliminary 2010 data. By 2008, some 1.3 billion people (22% of the developing world's population) were living on less than \$1.25 per day, down from nearly 2 billion (43%) in 1990, said the organization, in Washington DC. The trend has continued to 2010, although much of the fall is due to China's growing wealth.

THE DECLINE OF EXTREME POVERTY

The proportion of people living in extreme poverty (less than \$1.25 per day) in developing countries has fallen since 1990.



*\$1.25 marker is calculated in 2005 purchasing power parity dollars.

'Developing world' as defined by 1990 United Nations Millennium Development Goal.

NEWS IN FOCUS

PHYSICS Paradoxical particles could boost quantum computing **p.132**

LASER FUSION Nearing ignition, but will the technology catch fire? **p.133**

PUBLISHING Text-miners seek prospecting rights in journals **p.134**

JAPAN Taking stock of the tsunami disaster one year on **p.137**



Plans to provide drugs for people with HIV in Myanmar are being scaled back through lack of funds.

GLOBAL HEALTH

Cuts hamper bid to tackle AIDS

Medics warn of missed chance to use drugs to curb HIV.

BY ERIKA CHECK HAYDEN

Preventing the spread of HIV used to mean testing people for infection and encouraging them to practise safe sex. Increasingly, it also means prescribing drugs, as studies show that giving infected people or their uninfected partners antiretroviral drugs as soon as an infection is diagnosed can help to check the spread of AIDS.

Yet at this week's annual Conference on Retroviruses and Opportunistic Infections in Seattle, Washington, there was growing concern that financial austerity in the United States and elsewhere is eating away at the funding

needed for a worldwide prevention effort.

Many scientists and advocates agree that there is now an "awesome possibility to prevent the spread of HIV", says Sharonann Lynch, HIV policy adviser for Médecins Sans Frontières (MSF, also known as Doctors Without Borders) in New York. "If we decrease the money invested in treatment now, we are squandering the best opportunity we're going to have to get ahead of the wave of new infections."

Last month, US President Barack Obama's 2013 budget request proposed a 10.8% cut to

direct international aid for HIV programmes under the President's Emergency Plan for AIDS Relief (PEPFAR) which, together with previous cuts, would slice more than US\$1 billion from the fund's 2010 level (see 'Sliding support'). And last November, the Global Fund to Fight AIDS, Tuberculosis and Malaria said that it would not hand out any more funds for scaling up AIDS treatments until 2014 because of tightening budgets in donor countries (see *Nature* **480**, 159–160; 2011).

The shortfalls come as a slew of results presented this week reinforce a growing consensus about the power of early treatment for HIV infections. The latest data are part of a trend that accelerated last May, when HPTN 052, a clinical trial run by the multinational HIV Prevention Trials Network, showed that giving antiretroviral drugs to people who are HIV-positive can stop them from passing the virus to their uninfected partners (M. S. Cohen *et al.* *N. Engl. J. Med.* **365**, 493–505; 2011). In light of such results, the World Health Organization is expected to issue new guidelines for managing HIV in couples soon.

Other work discussed this week addressed some of the lingering questions over pre-exposure prophylaxis (PrEP), a strategy in which uninfected people take antiretroviral drugs to decrease their risk of contracting HIV. Epidemiologist Jared Baeten and biostatistician Deborah Donnell at the University of Washington in Seattle presented final data from the Partners PrEP study, one of three studies in the past 18 months to find that PrEP cut infection rates. The data support the study's initial findings and show that infection risk can be reduced by as much as 90% if the treatment is taken as prescribed.

That finding could account for a contrary result from FEM-PrEP, one of two studies that last year suggested PrEP is not effective for women. At the meeting, a principal investigator of FEM-PrEP, Lut Van Damme of the non-profit organization FHI 360 in Durham, North Carolina, reported that fewer than 40% of women in the study took their medication as intended.

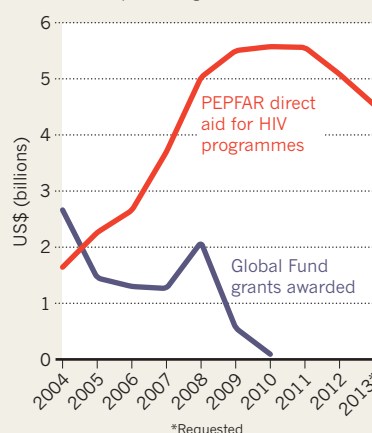
But many developing countries that depend on international aid to fight the AIDS epidemic may not be able to take advantage of these results. Funding cuts are already eroding existing treatment programmes, making any expansion unlikely.

In Myanmar, for instance, MSF has reported that fewer than 30,000 of the 120,000 people ▶

► **NATURE.COM**
For more, see
Nature's Outlook
on HIV/AIDS:
go.nature.com/iltuxi

SLIDING SUPPORT

Money available to two HIV funds has been cut back, compromising treatment efforts.



► who need antiretroviral drugs are getting them. The country had hoped to treat 46,500 more with money from the Global Fund, but abandoned that plan when the funding round was cancelled.

The Democratic Republic of Congo, which gives antiretrovirals to around 50,000 people with HIV — just 15% of those in the country who need them — is planning to lower its 2014 treatment targets from 82,000 to 54,000. It has also directed some non-governmental organizations to stop testing people for HIV, MSF says, because it cannot afford to treat people with the virus.

“Ultimately we’re going to have to wait for the Global Fund and PEPFAR to increase their resources if we are to truly impact this epidemic,” says Salim Abdool Karim, director of the Centre for the AIDS Programme of Research in South Africa, based in Durban. “We’re not going to be able to do it with the resources currently being made available.”

PEPFAR’s 2013 budget won’t be decided for many months. But Lynch and MSF are pushing the Global Fund to hold an emergency meeting, perhaps at June’s G20 summit in Los Cabos, Mexico, to drum up new money so that it can issue grants for HIV projects before 2014. The fund’s general manager, Gabriel Jaramillo, supports the idea.

Researchers, meanwhile, are finding ways to make do with less. Craig Cohen, an obstetrician and gynaecologist at the University of California, San Francisco, runs a programme that treats more than 114,000 people with HIV at 116 clinics across Kenya. Through money-saving measures such as reducing checks on patients who are stable on their medication, Cohen says that he has extended the programme’s reach, even as his funding from PEPFAR has declined. But, he adds, “you can only push so far to become more efficient”. ■

PHYSICS

A solid case for Majorana fermions

First compelling evidence of self-annihilating entities in a semiconductor is a step forward for quantum computing.

BY EUGENIE SAMUEL REICH

A year before he mysteriously disappeared at sea, Italian physicist Ettore Majorana posed a puzzle for future researchers. Quantum pioneer Paul Dirac had in 1928 predicted the existence of antimatter — mirror particles that annihilate with their matter counterparts. All known fermions — particles with half-integer spin, such as electrons — obey Dirac’s rules, but in 1937 Majorana varied Dirac’s equation to predict a class of particle that is its own antiparticle: the Majorana fermion.

Whether anything in nature fits Majorana’s prediction has remained unclear. But Leo Kouwenhoven, a nanoscientist at Delft University of Technology in the Netherlands, unveiled the most promising evidence so far for the existence of Majorana fermions, at a meeting of the American Physical Society in Boston, Massachusetts, on 27 February. The result is not just a vindication for the vanished theorist: Majoranas might also turn out to be the ideal information carriers in certain schemes for quantum computing.

What Kouwenhoven’s group spotted are not real particles, but quasiparticles formed from the collective movement of electrons in a semiconducting nanowire. Other researchers have invoked Majoranas to explain various observed behaviours in solid materials, but Kouwenhoven’s data are the first to clearly

demonstrate a predicted Majorana signature. “What he showed is extremely compelling,” says Jason Alicea, a theoretical physicist at the University of California, Irvine.

The Delft group tested a 2010 proposal that a pair of Majorana fermions could form at the interface between a superconductor and a semiconducting nanowire in a magnetic field (R. M. Lutchyn *et al.* Preprint at <http://arxiv.org/abs/1002.4033>; 2010). Majoranas are electrically neutral, and the mass of free-floating electrical charge in the superconductor allows electrons and absences of electrons — known as holes — to form neutral entities at the interface with the nanowire. Kouwenhoven reported a peak in the conductance through the nanowire at zero voltage: a signature of a spatially separated pair of Majoranas forming.

Patrick Lee, a physicist at the Massachusetts Institute of Technology in Cambridge, says that he wouldn’t call the observation a discovery just yet. He argues that other peaks in Kouwenhoven’s data point to the existence of non-Majorana states that could be mimicking the long-sought-after phenomenon. Even if Majorana fermions have surfaced, those other states could prove a problem for any possible application in quantum computing.

Normal particles are indistinguishable from others of the same type, so can swap positions in a system without a change in underlying quantum state. Majoranas, however, ‘remember’ swapping places, and an exchange leads to a new quantum state. This makes it possible to perform logical operations by physically exchanging Majoranas — ideal for quantum computing. “They’re very useful. A whole programme of research will follow from this,” says Charles Marcus, a physicist at Harvard University in Cambridge.

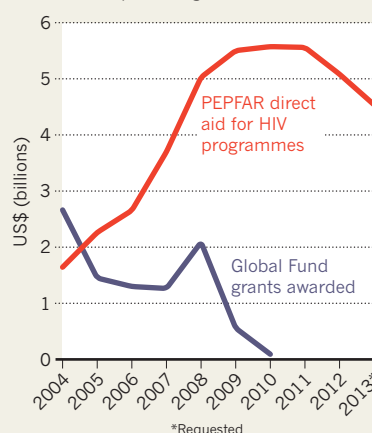
Physicists expect Majoranas to be created in spatially separated pairs that have a long-distance quantum link, as in Kouwenhoven’s experiment. The link should make the particles robust to decoherence — the destruction of quantum bits through interaction with the environment that dogs other quantum-computing approaches — says Roman Lutchyn, a theoretical physicist at Microsoft’s Station Q in Santa Barbara, California, and one of the authors of the proposal that Kouwenhoven tested. “This is a great development,” he says. ■



Ettore Majorana predicted non-Dirac antiparticles.

SLIDING SUPPORT

Money available to two HIV funds has been cut back, compromising treatment efforts.



► who need antiretroviral drugs are getting them. The country had hoped to treat 46,500 more with money from the Global Fund, but abandoned that plan when the funding round was cancelled.

The Democratic Republic of Congo, which gives antiretrovirals to around 50,000 people with HIV — just 15% of those in the country who need them — is planning to lower its 2014 treatment targets from 82,000 to 54,000. It has also directed some non-governmental organizations to stop testing people for HIV, MSF says, because it cannot afford to treat people with the virus.

“Ultimately we’re going to have to wait for the Global Fund and PEPFAR to increase their resources if we are to truly impact this epidemic,” says Salim Abdool Karim, director of the Centre for the AIDS Programme of Research in South Africa, based in Durban. “We’re not going to be able to do it with the resources currently being made available.”

PEPFAR’s 2013 budget won’t be decided for many months. But Lynch and MSF are pushing the Global Fund to hold an emergency meeting, perhaps at June’s G20 summit in Los Cabos, Mexico, to drum up new money so that it can issue grants for HIV projects before 2014. The fund’s general manager, Gabriel Jaramillo, supports the idea.

Researchers, meanwhile, are finding ways to make do with less. Craig Cohen, an obstetrician and gynaecologist at the University of California, San Francisco, runs a programme that treats more than 114,000 people with HIV at 116 clinics across Kenya. Through money-saving measures such as reducing checks on patients who are stable on their medication, Cohen says that he has extended the programme’s reach, even as his funding from PEPFAR has declined. But, he adds, “you can only push so far to become more efficient”. ■

PHYSICS

A solid case for Majorana fermions

First compelling evidence of self-annihilating entities in a semiconductor is a step forward for quantum computing.

BY EUGENIE SAMUEL REICH

A year before he mysteriously disappeared at sea, Italian physicist Ettore Majorana posed a puzzle for future researchers. Quantum pioneer Paul Dirac had in 1928 predicted the existence of antimatter — mirror particles that annihilate with their matter counterparts. All known fermions — particles with half-integer spin, such as electrons — obey Dirac’s rules, but in 1937 Majorana varied Dirac’s equation to predict a class of particle that is its own antiparticle: the Majorana fermion.

Whether anything in nature fits Majorana’s prediction has remained unclear. But Leo Kouwenhoven, a nanoscientist at Delft University of Technology in the Netherlands, unveiled the most promising evidence so far for the existence of Majorana fermions, at a meeting of the American Physical Society in Boston, Massachusetts, on 27 February. The result is not just a vindication for the vanished theorist: Majoranas might also turn out to be the ideal information carriers in certain schemes for quantum computing.

What Kouwenhoven’s group spotted are not real particles, but quasiparticles formed from the collective movement of electrons in a semiconducting nanowire. Other researchers have invoked Majoranas to explain various observed behaviours in solid materials, but Kouwenhoven’s data are the first to clearly

demonstrate a predicted Majorana signature. “What he showed is extremely compelling,” says Jason Alicea, a theoretical physicist at the University of California, Irvine.

The Delft group tested a 2010 proposal that a pair of Majorana fermions could form at the interface between a superconductor and a semiconducting nanowire in a magnetic field (R. M. Lutchyn *et al.* Preprint at <http://arxiv.org/abs/1002.4033>; 2010). Majoranas are electrically neutral, and the mass of free-floating electrical charge in the superconductor allows electrons and absences of electrons — known as holes — to form neutral entities at the interface with the nanowire. Kouwenhoven reported a peak in the conductance through the nanowire at zero voltage: a signature of a spatially separated pair of Majoranas forming.

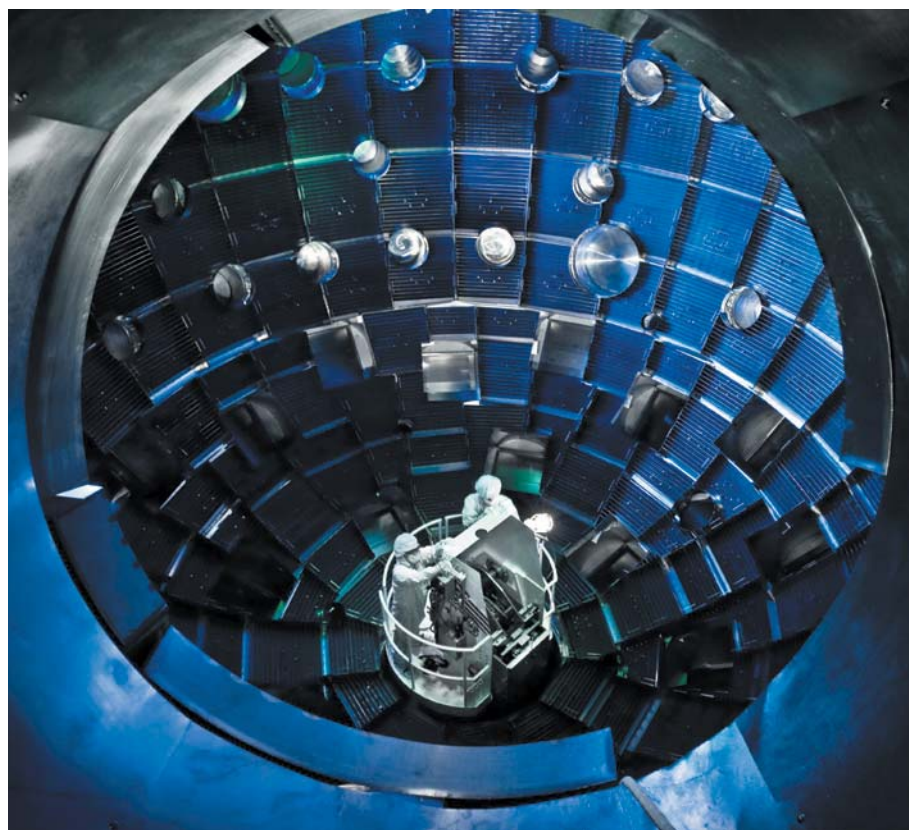
Patrick Lee, a physicist at the Massachusetts Institute of Technology in Cambridge, says that he wouldn’t call the observation a discovery just yet. He argues that other peaks in Kouwenhoven’s data point to the existence of non-Majorana states that could be mimicking the long-sought-after phenomenon. Even if Majorana fermions have surfaced, those other states could prove a problem for any possible application in quantum computing.

Normal particles are indistinguishable from others of the same type, so can swap positions in a system without a change in underlying quantum state. Majoranas, however, ‘remember’ swapping places, and an exchange leads to a new quantum state. This makes it possible to perform logical operations by physically exchanging Majoranas — ideal for quantum computing. “They’re very useful. A whole programme of research will follow from this,” says Charles Marcus, a physicist at Harvard University in Cambridge.

Physicists expect Majoranas to be created in spatially separated pairs that have a long-distance quantum link, as in Kouwenhoven’s experiment. The link should make the particles robust to decoherence — the destruction of quantum bits through interaction with the environment that dogs other quantum-computing approaches — says Roman Lutchyn, a theoretical physicist at Microsoft’s Station Q in Santa Barbara, California, and one of the authors of the proposal that Kouwenhoven tested. “This is a great development,” he says. ■



Ettore Majorana predicted non-Dirac antiparticles.



Engineers inspect the fusion chamber at the National Ignition Facility.

ENERGY

Laser fusion nears crucial milestone

National Ignition Facility approaches energy break-even point, but uncertainty over next step persists.

BY ERIC HAND

This could be the year the National Ignition Facility (NIF) finally lives up to its name. The facility, which boasts the world's largest laser, is designed to trigger fusion by imploding a target pellet of hydrogen isotopes, thereby releasing more energy than will go into the shot. NIF's managers think that the end of their two-year campaign for break-even energy, or 'ignition', is in sight. "We have all the capability to make it happen in fiscal year 2012," says Ed Moses, director of the US\$3.5-billion facility, at the Lawrence Livermore National Laboratory in California.

But even if the champagne corks do get popped, the method — a form of 'inertial confinement' fusion — faces an uncertain future. Would success mean that the US Department

of Energy (DOE) will be ready to develop it into an economically viable energy source? And if so, is NIF's laser-based approach the best one? An interim report released on 7 March by a US National Academies panel concludes that it is still too early to tell, and recommends that fusion scientists explore alternative technologies for imploding the fuel.

Glen Wurden, a plasma physicist at Los Alamos National Laboratory in New Mexico, agrees, saying that scientists working on inertial confinement should be wary of putting all their eggs in the laser basket. "It's premature right now," he says. He points to the troubles that have plagued a competing approach to fusion — magnetic confinement — and its flagship project ITER, a \$21-billion international fusion experiment under construction at St-Paul-lez-Durance, France. Wurden blames

ITER's delays and ballooning costs on a premature commitment to a technology known as a tokamak, a doughnut-shaped cage within which powerful electromagnets confine a fusion plasma.

Despite early confidence, bolstered by favourable computer models, NIF too has lagged behind schedule. "It thought it had ignition in the bag," says Wurden. Instead, NIF's approach to heating and compressing the hydrogen isotopes has proved troublesome. In what is known as indirect drive, the laser's multiple beams are focused at the openings in a pencil-eraser-sized gold cylinder called a hohlraum, blasting the insides to create X-rays. The X-rays then heat and squeeze the fuel pellet inside the hohlraum to produce fusion. But unexpectedly turbulent interactions between the laser light and the plasma inside the hohlraum sap energy from the beams. That could wipe out any gains as NIF managers ramp up the laser energy to the threshold needed for ignition.

The NIF team has made steady progress, however. When the push for ignition began 18 months ago, the facility was achieving 1% of the conditions thought to be needed for ignition. Now the figure stands at 10%, and the pace is quickening: a record 57 shots were taken in January alone (see 'Power play'). The team is also studying an array of tweaks, including encasing the fuel in beryllium or diamond instead of plastic and changing the hohlraum material or its shape. Moses says that it might also be possible to crank up NIF's peak energy from the 1.8 megajoules estimated to be needed for break-even to 2.2 megajoules.

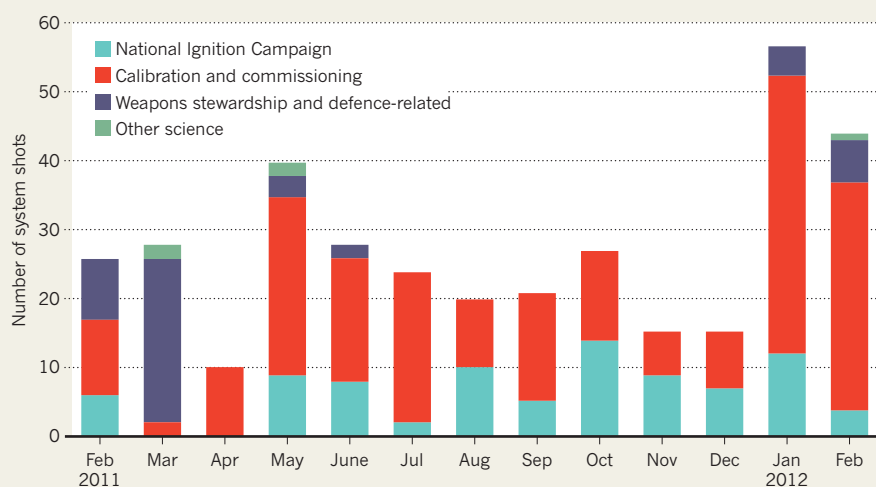
Still, as the National Academies' report notes, other approaches might provide an easier route to ignition and ultimately to a practical power plant. But who will pay for developing them? Most of the research on inertial confinement in the United States and worldwide has been supported by national security and weapons complexes, which want to study fusion for weapons purposes, not for electric power. Today, laser fusion in the United States makes its home within the National Nuclear Security Administration, the branch of the DOE responsible for stewardship of the nuclear stockpile.

Within the DOE's Office of Science, almost no money goes to inertial confinement research. The vast majority supports magnetic confinement fusion, and increasingly, that money is paying for ITER. Stephen Dean, president of Fusion Power Associates, an advocacy group based in Gaithersburg, Maryland, says that even if the academies' panel in its final report calls for a robust inertial confinement energy programme, the research will struggle to find a home with the office of science. "I think it'll just ignore it," he says. "ITER is obviously its top priority, and they're struggling like mad to save it."

The \$460 million requested for the NIF effort in 2013 will allow it to vary its ►

POWER PLAY

The National Ignition Facility's 192-beam laser achieved a record number of shots in January. Most were not directly part of the effort to ignite fusion.



► approach. For example, plasma physicists at the University of Rochester in New York want to adapt NIF's lasers so that they can implode a hydrogen-isotope pellet directly and dispense with the hohlraum.

The NIF scientists aren't waiting for

alternative approaches to catch up, however. Even before achieving ignition, they are racing to plan their next project, a demonstration power plant that they call LIFE, for Laser Inertial Fusion Energy. To be economic, the plant would have to produce more than 50 times

more energy from each shot than it puts in, and would have to boost repetition rates from a few shots a day to 15 per second — no mean feat.

In the quiet, cavernous NIF facility is a mock-up of one of the modular beam lines that would make up LIFE, small enough to fit in the back of a truck. Whereas NIF's set-up uses thousands of bulky flashbulbs to pump energy into the glass lasers, LIFE would use small, transistor-powered light-emitting diodes. Moses dismisses the notion that it's too early to commit to lasers as the drivers of a future power plant. Because of investment in lasers and transistors for consumer electronics, the world has already chosen, he says. Historians will look back, and "they'll see the transistor and the laser as the turning point".

LIFE director Mike Dunne says that the capital costs for the pilot plant would be about \$4 billion, and it could be putting hundreds of megawatts into the grid by the early 2020s — at least a decade earlier than the magnetic-fusion community hopes to deliver a practical power plant. Recalling the first time he presented the LIFE concept to magnetic-fusion researchers at a conference a few years ago, Moses says "The response to it was almost violent: 'This cannot be.' They were shocked at the ambition of it for sure. And they still are." ■

PUBLISHING

Trouble at the text mine

Computers can rapidly scan through thousands of research papers to make useful connections, but work is being slowed by publishers' unease.

BY RICHARD VAN NOORDEN

When he was a keen young biology graduate student in 2006, Max Haeussler wrote a computer program that would scan, or 'crawl', plain text and pull out any DNA sequences. To test his invention, the naive text-miner downloaded around 20,000 research papers that his institution had paid to access — and promptly found his IP address blocked by the papers' publisher.

It was not until 2009 that Haeussler, then at the University of Manchester, UK, and now at the University of California, Santa Cruz, returned to the project in earnest. He had come to realize that standard site licences do not permit systematic downloads, because publishers fear wholesale theft of their content. So Haeussler began asking for licensing terms to crawl and text-mine articles. His goal was to serve science: his program is a key part of the text2genome project, which aims to use DNA sequences in research papers to link the

publications to an online record of the human genome. This could produce an annotated genome map linked to millions of research articles, so that biologists browsing a genomic region could immediately click through to any relevant papers.

But Haeussler and his text2genome colleague Casey Bergman, a genomicist at the University of Manchester, have spent more than two years trying to agree terms with publishers — and often being ignored or rebuffed. "We've learned it's a long, hard road with every journal," says Bergman.

Many publishers say that they will allow their subscribers to text-mine, subject to contract and the text-miners' intentions, and point to a number of successful agreements. But like many early advocates of the technology, Haeussler and Bergman complain that publishers are failing to cope with requests, and so are holding up the progress of research. What is

more, they point out, as text-mining expands, it will be impractical for individual academic teams to spend years each working out bilateral agreements with every publisher.

With his frustration boiling over, Haeussler last week started a project to e-mail all the main science publishers for permission to mine their content. He will log their responses online (at <http://text.soe.ucsc.edu>) in the hope of raising awareness of the problem.

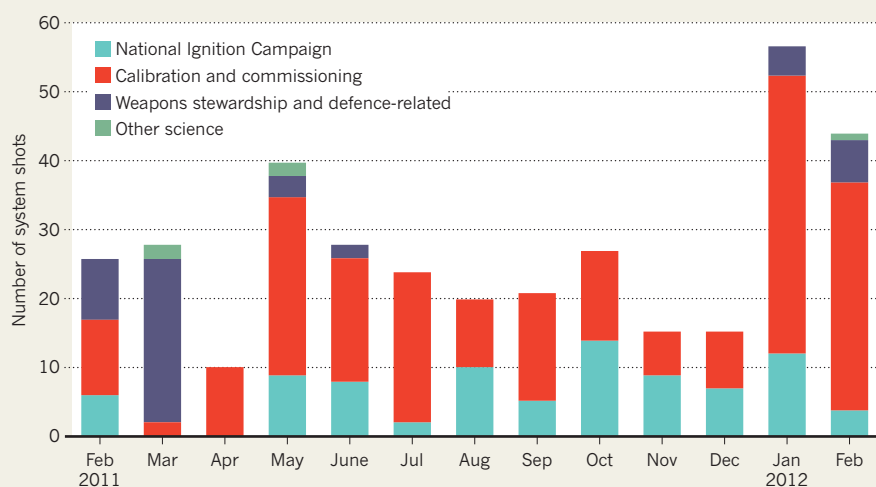
Academia is abuzz with excitement over text-mining. Thanks to growing computer power, software can recognize, extract and index scientific information from vast amounts of plain text, allowing computers to read and organize a body of knowledge that is expanding too fast for any human to keep up. 'Semantic software' is starting to record the relationships between scientific 'entities' — for example, between a particular drug and a specific enzyme.

For pharmaceutical firms, text-mining is "a basic necessity" that assists drug development, says Raul Rodriguez-Esteban, a computational

► **NATURE.COM**
To read more about
text-mining, see:
go.nature.com/rubywp

POWER PLAY

The National Ignition Facility's 192-beam laser achieved a record number of shots in January. Most were not directly part of the effort to ignite fusion.



► approach. For example, plasma physicists at the University of Rochester in New York want to adapt NIF's lasers so that they can implode a hydrogen-isotope pellet directly and dispense with the hohlraum.

The NIF scientists aren't waiting for

alternative approaches to catch up, however. Even before achieving ignition, they are racing to plan their next project, a demonstration power plant that they call LIFE, for Laser Inertial Fusion Energy. To be economic, the plant would have to produce more than 50 times

more energy from each shot than it puts in, and would have to boost repetition rates from a few shots a day to 15 per second — no mean feat.

In the quiet, cavernous NIF facility is a mock-up of one of the modular beam lines that would make up LIFE, small enough to fit in the back of a truck. Whereas NIF's set-up uses thousands of bulky flashbulbs to pump energy into the glass lasers, LIFE would use small, transistor-powered light-emitting diodes. Moses dismisses the notion that it's too early to commit to lasers as the drivers of a future power plant. Because of investment in lasers and transistors for consumer electronics, the world has already chosen, he says. Historians will look back, and "they'll see the transistor and the laser as the turning point".

LIFE director Mike Dunne says that the capital costs for the pilot plant would be about \$4 billion, and it could be putting hundreds of megawatts into the grid by the early 2020s — at least a decade earlier than the magnetic-fusion community hopes to deliver a practical power plant. Recalling the first time he presented the LIFE concept to magnetic-fusion researchers at a conference a few years ago, Moses says "The response to it was almost violent: 'This cannot be.' They were shocked at the ambition of it for sure. And they still are." ■

PUBLISHING

Trouble at the text mine

Computers can rapidly scan through thousands of research papers to make useful connections, but work is being slowed by publishers' unease.

BY RICHARD VAN NOORDEN

When he was a keen young biology graduate student in 2006, Max Haeussler wrote a computer program that would scan, or 'crawl', plain text and pull out any DNA sequences. To test his invention, the naive text-miner downloaded around 20,000 research papers that his institution had paid to access — and promptly found his IP address blocked by the papers' publisher.

It was not until 2009 that Haeussler, then at the University of Manchester, UK, and now at the University of California, Santa Cruz, returned to the project in earnest. He had come to realize that standard site licences do not permit systematic downloads, because publishers fear wholesale theft of their content. So Haeussler began asking for licensing terms to crawl and text-mine articles. His goal was to serve science: his program is a key part of the text2genome project, which aims to use DNA sequences in research papers to link the

publications to an online record of the human genome. This could produce an annotated genome map linked to millions of research articles, so that biologists browsing a genomic region could immediately click through to any relevant papers.

But Haeussler and his text2genome colleague Casey Bergman, a genomicist at the University of Manchester, have spent more than two years trying to agree terms with publishers — and often being ignored or rebuffed. "We've learned it's a long, hard road with every journal," says Bergman.

Many publishers say that they will allow their subscribers to text-mine, subject to contract and the text-miners' intentions, and point to a number of successful agreements. But like many early advocates of the technology, Haeussler and Bergman complain that publishers are failing to cope with requests, and so are holding up the progress of research. What is

more, they point out, as text-mining expands, it will be impractical for individual academic teams to spend years each working out bilateral agreements with every publisher.

With his frustration boiling over, Haeussler last week started a project to e-mail all the main science publishers for permission to mine their content. He will log their responses online (at <http://text.soe.ucsc.edu>) in the hope of raising awareness of the problem.

Academia is abuzz with excitement over text-mining. Thanks to growing computer power, software can recognize, extract and index scientific information from vast amounts of plain text, allowing computers to read and organize a body of knowledge that is expanding too fast for any human to keep up. 'Semantic software' is starting to record the relationships between scientific 'entities' — for example, between a particular drug and a specific enzyme.

For pharmaceutical firms, text-mining is "a basic necessity" that assists drug development, says Raul Rodriguez-Esteban, a computational

► **NATURE.COM**
To read more about
text-mining, see:
go.nature.com/rubywp

biologist at the drug giant Boehringer Ingelheim in Ridgefield, Connecticut. Companies routinely create custom databases of proteins, drugs, cell types and the interactions between them, all gleaned from text-mining, he explains. The technology still needs human oversight, but most enthusiasts expect text-mining to be the key to a new kind of scientific discovery based on rich, computer-readable representations of knowledge gathered from plain-text research articles.

But, as Haeussler has discovered, there is a major roadblock. Freely available patents and article abstracts are open for text-mining, but material behind paywalls is not — even when institutions have paid for a site licence. “The licence is oriented towards permitting the human to download and read an article, but not to text-mine it,” says John McNaught, deputy director of the National Centre for Text Mining at the University of Manchester. Even freely accessible papers may not come with permissive licences: of the 2.4 million abstracts listed by PubMedCentral, only 400,000 (17%) are licensed for text-mining.

ILLICIT PROSPECTING

Software programmers can circumvent publishers’ detection systems, for example by ensuring that papers aren’t crawled or downloaded in one batch. This breaches the normal site licence terms, but Haeussler says that papers derived from such technically illegal text-mining have been published in leading journals.

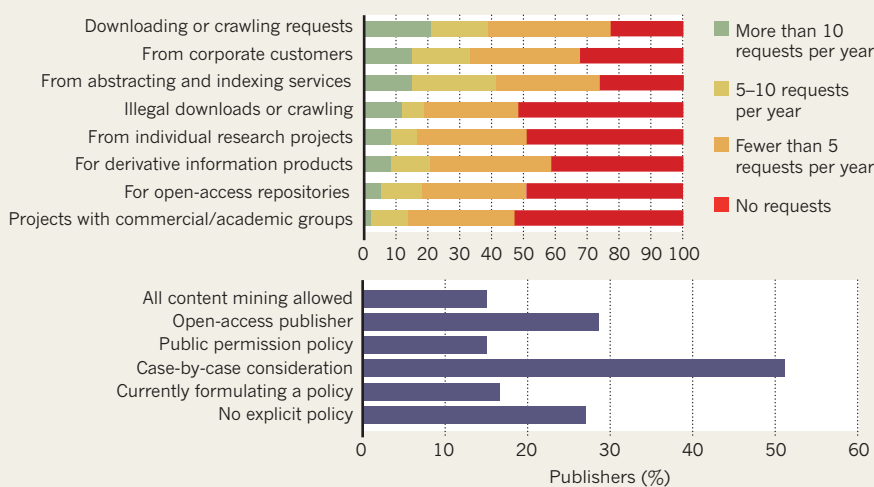
Those wishing to text-mine within the rules must agree contracts with the publishers, and sometimes pay a fee. Haeussler got permission to mine the corpus of Amsterdam-based publisher Elsevier for free. But another academic text-mining project, BioNOT, based at the University of Wisconsin–Madison, was not so fortunate. The collaboration was charged extra for its contract to search Elsevier papers to automatically extract negative results, potentially useful for showing that genes are not related to a disease, for example.

Even powerful drug firms find the negotiations a burden. “When we have licensed and paid for the full text, we feel that we should also have the right to mine it,” says Henning Nielsen, head of the Library and Information Centre at the Danish pharmaceutical firm Novo Nordisk in Bagsværd, Denmark, and president of the Pharma Documentation Ring (PDR), an association of information managers covering 21 of the world’s largest drug firms.

Publishers deal with text-mining requests in various ways. Last year, the Publishing Research Consortium (PRC), a trade body that supports research on scholarly communication,

MINE ALL MINE

About 75% of the 190 publishers polled by the Publishing Research Commission have received text-mining requests (top), but they have a range of policies (some have more than one) on granting access (bottom).



commissioned a survey about content-mining, for which it polled 190 journal publishers (E. Smit and M. Van der Graaf *Learn. Publ.* 25, 35–46; 2012). Of these, 48% said that they had detected illegal crawling and downloads of their content, and 51% had received requests from individual research projects — although most had received fewer than five requests per year (see ‘Mine all mine’). More than half of publishers said that they decide on a case-by-case basis whether to allow access. Of these, one-third said that they would charge for it if the request was for commercial purposes. For example, some publishers seem concerned that if someone text-mines their content to produce a marketable product, it could compete with or supplant their own content. Nature Publishing Group in London, which publishes this journal, says that it does not charge existing subscribers to mine content to which they already have access, subject to contract.

There are signs that policies may soon be clarified. Nielsen says that the PDR hopes to hammer out a solution with major publishers this year, to allow drug firms to text-mine the literature more easily. And last August, the UK government accepted the recommendations of an intellectual-property review that said scientists should be allowed to mine text and data from journal articles without having to ask permission from a copyright owner — although this has not become law, and does not trump current licence agreements, which tend to bar systematic downloading of papers.

On 8 March, the Copyright Clearance Center — an organization based in Danvers, Massachusetts, that works with publishers on rights licensing — is holding a forum in Amsterdam to discuss what publishers should do about text-mining. And the International Association of Scientific Technical and Medical Publishers, a trade body based in Oxford, UK, says that it is working to agree a shared

position on text and data mining, which it expects to resolve by the summer.

Increasingly, publishers are starting to recognize the opportunities of text-mining, and to mine their own content. The PRC survey found that just under half of publishers said that they already do so, with almost one-third of the rest planning to start this year. The work — often contracted out to the same third-party text-mining firms that are employed by the pharmaceutical industry — typically involves computer programs picking out all the chemicals, genes or proteins from a research paper, and in some cases uploading them to online databases.

LIMITED ACCESS

Elsevier is now actively inviting text-miners, including BioNOT, to write programs (or ‘apps’) that crawl through the full text of its research articles to pick out information. Subscribers to Elsevier’s website can access more than 100 of these apps — including Haeussler’s program. But the apps run only within the website, and contracts usually stipulate that the mined content cannot be used elsewhere. This, says Bergman, is of limited use, because the publisher covers only a small amount of the research literature. He and others shudder at the prospect of individual publishers making text-miners adopt different standards, or stipulating that a particular text-mining program can be used only on their papers — effectively destroying the technology’s potential to crawl across the entire research literature.

Publishers are still working out how to take advantage of text-mining, but none wants to miss out on the potential commercial value. “The technology is progressing so quickly that publishers haven’t had time to think it through,” says David Haussler of the University of California, Santa Cruz, who leads the text2genome project. “As soon as they do, they will realize this is a wonderful opportunity.” ■ [SEE EDITORIAL P.124](#)



AFTER SHOCK

Eighteen hours after a massive earthquake ripped apart the sea floor off the coast of Tohoku, Japan, on 11 March last year, tsunami waves smashed the coast of Antarctica hard enough to knock loose giant icebergs. Over the next few weeks, as an estimated 19,000 people were reported dead or missing in Japan, radioactive isotopes from the crippled Fukushima Daiichi nuclear power plant circled the globe. The earthquake, tsunami and extended nuclear crisis that devastated Japan affected the entire planet and continue to have global repercussions. This week, *Nature* dissects what has been learned about the event over the past year and how those lessons can help societies to prepare for future disasters.

Efforts to rebuild the Tohoku coastline have provoked debate about how to best protect Japan from future tsunamis through zoning and physical defences (**page 141**). Meanwhile, Japan and the United States are upgrading their systems of sensors that forecast these destructive waves (**page 144**). Seismologist Hiroo Kanamori says that tsunami forecasters should improve their ability to analyse both the magnitude and mechanism of an ongoing earthquake (**page 147**). And

geologist Thorne Lay examines how a spate of giant earthquakes in the past decade has revealed gaps in seismological knowledge (**page 149**).

Researchers are accumulating a vast amount of data on the causes and impacts of the Fukushima disaster — but some worry that their efforts are being undermined by a public loss of trust in the government and its scientists (**page 138**). And energy-policy expert Peter Bradford argues that countries should develop sensible policies towards nuclear power, rather than shunning it in a rash reaction to the Fukushima meltdowns (**page 151**).

Scientists have had a key role in helping Japan to recover and gird itself for the next disaster. But even this nation — one of the most at risk for earthquakes and tsunamis, and so one of the best prepared — has far to go. Other countries would do well to heed the lessons. ■

Devastation followed the earthquake in Japan last March.

PAOLO PELLEGRINI/MAGNUM PHOTOS



A YEAR AFTER THE TSUNAMI

For more content, go to:
www.nature.com/japanquake

FUKUSHIMA'S LEGACY OF FEAR

BY GEOFF BRUMFIEL
AND ICHIKO FUYUNO

Japan's worst-ever nuclear accident displaced more than 100,000 people. Many could now safely return home. Yet mistrust of the government prolongs their exile.

Yoichi Tao is busily shovelling dirt in Iitate, a small village about 40 kilometres from the ruined Fukushima Daiichi nuclear plant.

It is certainly different from his day job. Tao has a background in high-energy physics, and teaches students about information-security systems at Kogakuin University in Tokyo. But on this sunny February morning, he and a dozen volunteers have joined local farmers in removing the top few centimetres of radioactive soil from rice fields. "When the soil is frozen, we can remove it easily, like a board," Tao says. In a corner of the field, they dump the soil in a hole lined with absorbent sheets. "When spring comes, the ice melts but the [radioactive] caesium will be absorbed, so we can protect it from leaking out," he explains.

The volunteers, mostly researchers, informally call themselves *Fukushima Saisei-no Kai* (roughly translated as the Fukushima revitalization association), and come armed with car-mounted sodium iodide scintillators and Geiger counters linked to the Global Positioning System. Only in one of the world's most technically advanced societies could an ad hoc group have the means to cope with radioactive decontamination.

But there is a dark side to Tao's efforts: he is there because he and many others have lost faith in their government. "Since 11 March, people haven't trusted scientists who receive

funding from the government," Tao says. "They trust people who act without government funding and who work together with them."

One year after Japan's nuclear crisis began, researchers contacted by *Nature* say that a strong, evidence-based understanding of the accident, and the risks the reactors continue to pose, is within reach. The findings could inform decisions on public health, environmental clean-up and economic recovery (see 'The fallout'). But outside observers, and even some critics in Japan, are increasingly worried that the loss of public trust, together with politicians' desperation to regain it, could undermine rational decision-making about clean-up and resettlement. At stake are the futures of more than 100,000 residents who have been displaced from the area around the plant, and billions of dollars in economic activity across the region.

MELTDOWN

The crisis began on 11 March 2011, when a magnitude-9.0 earthquake on the Pacific floor sent a massive wall of water rolling towards the Japanese coastline (see page 141). The three operating reactors at Fukushima Daiichi automatically shut down in the moments after the

quake, but 41 minutes later the tsunami burst through the plant's defences and inundated the reactor buildings. Water flooded emergency generators, leaving the plant without power for cooling systems, while radioactive decay continued to heat the cores. In the control room, workers struggled to run crucial instruments, using torches and car batteries scavenged from nearby vehicles. Over the following days, the last line of emergency systems failed and the three reactors melted down. The process released hydrogen gas, which eventually triggered explosions in the reactor buildings. Volatile radioactive chemicals, notably iodine-131 and caesium-137, began to stream into the air and sea.

When unit 4 of the Chernobyl nuclear power plant exploded in 1986, the Soviet government imposed a strict information blackout. The situation could hardly have been more different at Fukushima: within the first 24 hours, the government began reporting radiation readings. In the following days and weeks, the deluge of information became swollen with data from university researchers, the military, international monitors, representatives of the US government and concerned citizens such as Tao.

"We've almost got too much," says Malcolm Crick, secretary of the United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) in Vienna. At the UN's request, the committee has spent the past six months



A YEAR AFTER THE TSUNAMI

For more content, go to:
www.nature.com/japanquake

trying to unpick which data came from where, and how they were calibrated. The committee will deliver its preliminary findings in May, and Crick says they should be able to say a great deal about how much radioactivity was released, where it went, and how much workers and the general public received.

DOSE LEVELS

It is already evident that rapid evacuation and careful screening protected Fukushima's citizens from harm, says Wolfgang Weiss, a physicist at Germany's Federal Office for Radiation Protection in Munich and chair of UNSCEAR. Early and informal analyses by his colleagues suggest that no members of the public received a dangerous dose of radiation.

That finding is supported by a sweeping public-health study begun last summer at Fukushima Medical University. With a ¥78.2-billion (US\$958-million) budget, the survey is designed to monitor the health of some 2 million people from the region for 30 years. According to the latest estimates, released on 20 February, 99.3% of 9,747 people living in towns or villages close to the plant received less than 10 millisieverts (mSv) in accumulated effective dose in the first four months after the accident. The highest recorded dose was 23 mSv, well below the acute 100-mSv exposure levels linked to a slight increase in cancer risk.

Yet suspicion is hampering the ambitious health survey, which hopes to nail down the long-term impact of Fukushima on ordinary citizens. Despite efforts to promote the study among evacuees, participation stands at just 21%. "Most of the people I've met here refuse to fill in the questionnaires. They don't see credibility in what the government does, and they say, 'this is just a survey of guinea pigs,'" says Shizuko Otake of the non-profit organization Shalom, which supports refugees in neighbouring Minamisoma and Iitate.

The roots of mistrust can be traced to the confusing days immediately after the explosions, when authorities made a series of inconsistent statements, issuing radiation readings that often turned out to be incorrect. As radioisotopes spread from the plant, the government was repeatedly forced to raise its recommended safety limits for radiation exposure to citizens and workers — otherwise, it would have been legally required to evacuate the site immediately. As a result, some Japanese people believe that the government is corrupt; others think it is incompetent. The prevailing feeling is that "what the government says always changes", Otake says.

Abel González, a radiation-protection expert with Argentina's nuclear regulatory authority in Buenos Aires, says that the government was forced to raise the safe limits because it started with an international standard that made no provision for accident scenarios or for emergency workers likely to receive higher



Citizen clean-up crews carry out rudimentary decontamination in villages such as Minamisoma.

doses. Without clear guidelines, the Japanese government simply had to increase the safe limits to enable people to keep working to bring the nuclear plant under control, he says. The downside is that "when you relax the regime in the middle of an accident, you lose credibility immediately".

In an effort to win back the trust of its citizens, the government is planning one of the most extensive and costly clean-up operations ever — an effort some experts view as unrealistic. Last autumn, it announced plans to bring

**"SINCE 11 MARCH, PEOPLE
HAVEN'T TRUSTED SCIENTISTS
WHO RECEIVE FUNDING FROM
THE GOVERNMENT."**

radiation doses from the accident to below 1 mSv per year in as much of the evacuation zone as possible. But the goal is based on an international standard for doses received during the normal operation of a nuclear plant, not following an accident.

It is also seen by veterans of nuclear accidents as highly ambitious, especially given the mountainous and heavily wooded terrain around Fukushima. "The best thing to do, according to Chernobyl, is to really turn the first metre of soil upside down," says Weiss. "But if you do that, you would kill the whole ecosystem."

The Japanese authorities acknowledge the problem, and have started trialling a variety of clean-up methods in Fukushima. The most

prominent pilot project began last November under the Japan Atomic Energy Agency (JAEA), with an estimated budget of ¥10.9 billion. The JAEA contracted the project to joint ventures led by three major construction companies — Taisei, Obayashi and Kajima — which are testing various technologies to clean up radioactive materials in 11 cities, towns and villages whose citizens mostly remain evacuated.

"I am impressed how companies have come up with novel ideas to remove decontaminated caesium based on existing technologies," says Shinichi Nakayama, deputy director of the JAEA's Fukushima Environmental Safety Center. For example, scouring caesium from roads with a high-pressure water jet was thought to be insufficient because contaminated water would simply spread out across the pavement. But engineers have modified the system to recover the contaminated water, purifying and recycling it, he says.

Many communities are taking matters into their own hands. With the help of independent researchers like Tao, they are removing contaminated soil and conducting other clean-up operations. But without a central disposal location, Weiss says, these clean-up operations are just creating a different waste problem. "People are not allowed to transport the waste, so they put everything in holes on their property."

The government's ambitious goals for decontamination could harm evacuees by inciting needless fears, says Oleg Nasvit, a radioecologist at the National Institute for Strategic Studies in Kiev, Ukraine, who has studied the impact of the Chernobyl nuclear accident. In 1986, the Soviet authorities demanded

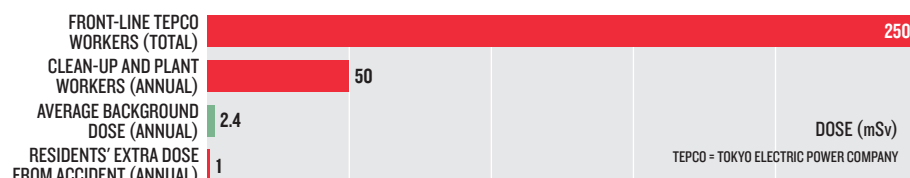
“obligatory evacuation” of residents living in regions where the additional radiation exposure from the accident was greater than 5 mSv per year, he says. Evacuees struggled to cope with their dislocation, and many were stigmatized because they had come from a contaminated region. “Frankly, this brought to people more harm than good,” he says.

Setting low radiation-dose limits is already damaging the economy around Fukushima. Later this year, the health ministry is planning to lower the safe level for caesium in vegetables, grain and other foods from 500 becquerels per kilogram (Bq kg⁻¹) to 100 Bq kg⁻¹ (see ‘The limits’). Tomoko Nakanishi, a researcher specializing in plant radiophysiology at the University of Tokyo, says that food with radioactivity lower than 500 Bq kg⁻¹ is not harmful to human health, and that areas not heavily affected by Fukushima’s radioisotopes may already exceed the proposed lower limits because of older nuclear fallout. Some mushrooms from Chiba prefecture, more than 200 kilometres south of Fukushima, exceed 100 Bq kg⁻¹, for example, but the relative amounts of radioisotopes are characteristic of residual contamination from nuclear weapons tests in the 1950s and 1960s or the Chernobyl accident, not Fukushima.

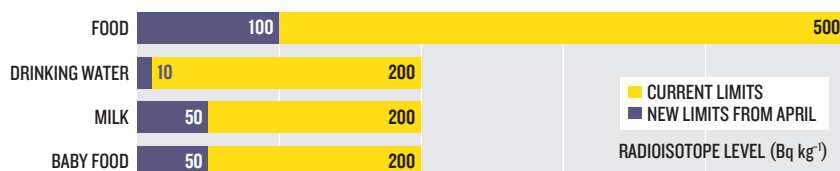
Fukushima prefecture is among the largest rice producers in Japan, but last year the agriculture ministry considered completely prohibiting cultivation where crops contained more than 100 Bq kg⁻¹ of caesium. Nakanishi and her colleagues at the University of Tokyo were concerned that the excessively stringent safety measures could hinder not only the recovery of the region’s agriculture, but also the collection of scientific data. “Continuous cultivation is very important to predict what will happen in the future. We don’t know if rice paddies that produced high-level caesium will do the same this year,” she says. Backed by strong demand from farmers, the ministry recently decided to allow cultivation in most areas of Fukushima as long as cities, villages and towns can prevent the distribution of rice containing more than 100 Bq kg⁻¹ of caesium.

THE LIMITS

The Japanese government wants citizens’ exposure to Fukushima’s radiation to be less than half their annual dose from natural sources.



To meet this goal, Japan plans to enforce stricter contamination limits for food and drink.



THE FALLOUT

Risks linger in reactors and the environment

The reactors In the immediate aftermath of the accident, operators could do little other than flush the overheated reactors with sea water using a fire engine as an improvised pump. Today, the situation at the plant has improved dramatically. Corrosive salt water has been replaced with fresh water that is passed through a filtration system to remove radioactive caesium before being recycled back into the cores.

The system seems to be working reasonably well. Late last year, temperatures at all three reactors dropped below 100°C, leading officials to announce that the plants had achieved ‘cold shutdown’. But dangers remain. The plant continues

to leak water, and an analysis by Japanese researchers warns that recent seismic activity could hint at a looming earthquake that might threaten the plant.

Meanwhile, Japan has developed a ‘stress test’ to assess the safety of its other nuclear plants. But for now, just two of the country’s 54 reactors are running.

The environment Fukushima has stopped releasing radioisotopes into the air, and radioactivity in the sea seems to have dispersed with little effect. But some organisms may still be accumulating nuclear material from the plant.

Researchers at the Forestry and Forest

Products Research Institute in Tsukuba have found earthworms containing nearly 20,000 becquerels per kilogram of radioactive caesium in Kawauchi, 26 kilometres from the plant. And local bird populations seem to have declined by about a third, according to Tim Mousseau, a radioecologist at the University of South Carolina in Columbia (A. P. Møller *et al. Environ. Pollut.* **164**, 36–39; 2012). In the ocean, says Ken Buesseler, a marine chemist at Woods Hole Oceanographic Institution in Massachusetts, radioactive plutonium and strontium seem to be finding their way into fish and bottom-feeders living near the reactor. **G.B. & I.F.**

low levels, a move that Nasvit says makes the plan more credible because the areas that will benefit the most from decontamination will be dealt with first.

At the moment, there is no clear plan for allowing displaced residents to go home. Although the government’s goal is to ensure that people should not receive a dose in excess of 1 mSv per year if they return, it is not a firm rule. Indeed, locals can already go home to villages outside the 20-kilometre restricted zone around the plant if they choose, but many public facilities, such as schools, have not yet reopened. Nasvit believes that citizens should move back, even to zones where they might receive up to 20 mSv per year. González agrees, noting that in some parts of the world, natural annual levels of radiation are in the range of 10–100 mSv.

But Tatsuhiko Kodama, director of the Radioisotope Center of the University of Tokyo, thinks the safety margin is not so clear cut. “There are various interpretations about what to do in the area of 1–20 mSv per year,” he says. He agrees, though, that ultimately the public must choose the course of action. “The most important thing is to respect what the residents think. We have to proceed with plans based on their decisions.”

Fred Mettler, a radiologist serving on the UNSCEAR panel, agrees. Rather than setting a strict number or limit, he says, the discussion should be more open ended. “We tell the people what’s there, we tell people what the consequences are, and they decide whether to accept the risk.” ■ **SEE EDITORIAL P.123**

AFTER THE DELUGE

JAPAN IS REBUILDING ITS COASTAL CITIES TO PROTECT PEOPLE FROM THE BIGGEST TSUNAMIS.

BY DAVID CYRANOSKI

On a cold February day, the northeastern Japanese city of Sendai is a snow-covered wasteland. It might pass for out-of-season farmland were it not for the chunks of grey concrete arrayed in rectangles, reminders that houses once stood here.

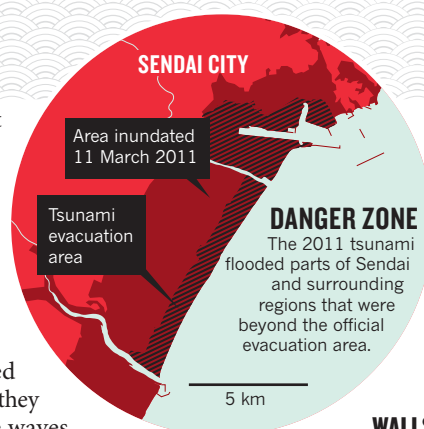
The signs of devastation are all around, although eerily tidied up. Inside one abandoned house that is missing its first-storey walls, dishes have been neatly stacked on a shelf, perhaps by a compassionate rescue worker. The remnants of cars — 240,000 were washed away or destroyed — and other metallic wreckage are compressed into neat blocks and stacked. Nearby lie piles of uprooted trees that were supposed to provide protection against a tsunami, but instead became lethal battering rams in the raging water. A solitary street sign lies on the ground by the beach, warning of the risk posed by such events.

The sign shows areas that must be evacuated in the event of a tsunami, but the stretch of land at risk on the map is just a small fraction of the region of Sendai that was flooded on 11 March 2011 (see 'Danger zone'). Fumihiko Imamura, a tsunami researcher at Sendai's Tohoku University, quietly studies the map, which he helped to produce. "Some people have criticized us, but

at that time we were just envisioning the Miyagi quake," he says, referring to an earthquake that occurs off the Sendai coast every 100 to 150 years, triggering tsunamis about 4 metres high. Instead, water levels here reached 10 metres; farther north they topped 20 metres. The waves inundated more than 500 square kilometres of land across six prefectures, destroying nearly 130,000 buildings and damaging 245,000 others. Some 15,000 people died, and several thousand more are missing and presumed dead.

Now the Japanese government is wrestling with how to rebuild cities such as Sendai to keep people safe should a monster tsunami hit again. Scientists, architects and city planners are debating how much to rely on coastal fortifications such as sea walls and forests. These protections helped in some places but many of them failed, and they gave people a false sense of security that may have contributed to the death toll. Despite this, many municipalities are already moving forward with reconstructing damaged defences.

A new institute, opening in Tohoku next month, will try to extract lessons from the disaster to aid the rebuilding efforts and to develop better public-education campaigns, which experts say are one of the most effective



tools for protecting people. Tatsuo Hirano, minister of the recently launched Reconstruction Agency, says that Japan is setting its sights high as it starts to rebuild. "The goal is to have zero deaths in future tsunamis," he says.

WALLS VERSUS WATER

Japan has been pounded by tsunamis throughout its history, and so had prepared more for these floods than any other country. Over the centuries, communities have planted coastal forests as barriers against the sea. Many towns hold regular training drills and evacuations. And sea walls and breakwaters surround nearly half of the country's 34,500 kilometres of coastline. Maintaining and extending these bulwarks has cost the central government billions of yen per year. In areas of greatest tsunami risk, including the entire Tohoku region, 74% of sea walls were built to be higher than the expected tsunamis.

But those expectations were based on magnitude-8 earthquakes that occur every few decades or centuries, not the 1-in-1,000-year Tohoku event, which was ten times stronger at magnitude 9.0. The tsunami that came crashed over most of those walls. Although the giant waves shocked the public, they weren't a complete surprise to some researchers. A decade ago, geologists discovered a sedimentary



A YEAR AFTER THE TSUNAMI

For more content, go to:
www.nature.com/japanquake

layer from a massive tsunami that had flowed several kilometres inland in 869 (K. Minoura *et al. J. Nat. Disaster Sci.* **23**, 83–88; 2001). Some scientists had worried about a repeat. “We thought one like this might come, but it was not yet in the official model,” says Imamura. During a meeting with Imamura last month, Tsuneaki Iguchi, the mayor of the city of Iwanuma south of Sendai, thanked the scientist for his efforts to protect the region but said he wished that researchers had done better at getting the message out about the 869 tsunami. “If only that work could have been done a little more quickly.”

Japan is now taking steps to shore up its defences against such mega-tsunamis. In December, the parliament passed a law requiring the construction of “tsunami-safe cities”. Sea walls and other structures are meant to provide complete protection from “the tsunami that comes every 200–300 years”, says Hirano. In the case of the biggest tsunamis, local governments will use zoning restrictions to prevent people from living in low-lying areas and will improve evacuation protocols to augment the protection provided by sea walls.

The debate over how to protect cities has sometimes been fierce. Some have questioned the value of expensive sea walls. According to Hirano, of 300 kilometres of such walls in the Tohoku region, 180 kilometres were swept away or destroyed after the Tohoku quake. That includes sections of a ¥120-billion (US\$1.5-billion) breakwater in Kamaishi Port,

which had been completed just three years earlier. The government announced last year that ¥55 billion would be allocated to repairing it.

Supporters say that the Kamaishi breakwater helped to reduce damage from the tsunami. According to simulations by the Port and Airport Research Institute in Yokosuka, the wall reduced the wave’s height at landfall from 13.7 metres to 8.0 metres, cut its maximum height inland by 50% and gave residents an extra 6 minutes to evacuate.

But those estimates may not be reliable, says Ioan Nistor, a coastal engineer specializing in tsunamis at the University of Ottawa in Canada, who visited Kamaishi just after last year’s quake. He says the analysis incorrectly assumed that the breakwater remained intact during the tsunami. “Although the presence of this breakwater has had a certain positive benefit,” he says, “given that a number of sections were knocked off by the tsunami impact, I am not sure if one can quantify so precisely its benefits.”

FAILED FORESTS

At the other end of the cost spectrum are coastal forests, says Kentaro Imai of Tohoku University. During the past few centuries, coastal cities all around Japan have planted them, especially those in the tsunami-prone Tohoku region. But of the 230 kilometres of protective coastal forests, two-thirds were heavily damaged by last year’s tsunami.

For the most part, the trees did more harm than good, says Hermann Fritz, who studies

the fluid dynamics of natural disasters at the Georgia Institute of Technology in Savannah. Fritz and some Japan-based colleagues carried out a survey of the city of Rikuzentakata, where the tsunami reached heights of 15 metres and destroyed a 200-metre-wide forest before heading inland and laying waste to large sections of the city. Only one tree — later named the ‘tree of hope’ — stood firm.

“There was no ‘tsunami control’ by the forest,” says Fritz. “It became 70,000 rams of floating debris impacting buildings.” He is not surprised. In a post-disaster survey of Tohoku’s Kesennuma Bay, his team measured flow velocities of about 10 metres per second for the water that coursed through the city of Kesennuma (H. M. Fritz *et al. J. Geophys. Res.* **39**, L00G23; 2012). “There is no way a forest will survive that,” Fritz says.

Yet the Japanese government has decided to invest ¥59 billion in replanting trees in Tohoku. Proponents argue that the trees also serve other purposes, such as providing a wind break that stops sand from blowing inland. And there is evidence that the forests have slowed tsunami waves resulting from some smaller quakes. Even last year there were some examples of success. In Hachinohe, which was hit by waves higher than 6 metres, the trees stood firm and blocked more than 20 boats from being swept inland and causing further destruction.

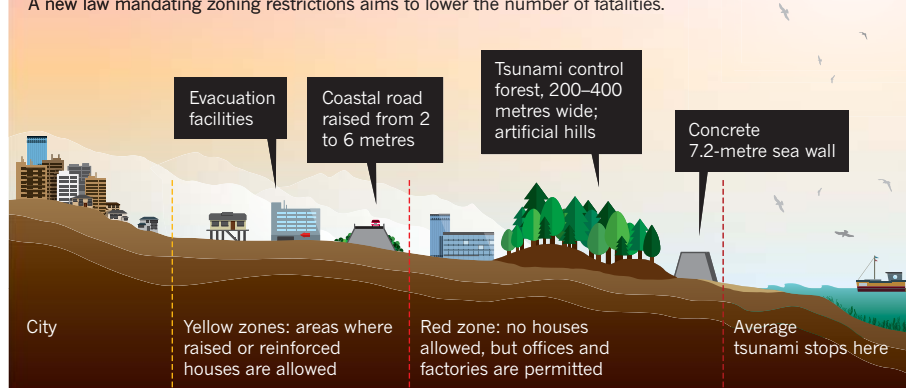
Researchers hope that clues from last year’s disaster can help them to improve the performance of coastal forests. Toyohiko Miyagi,



Raging waters breached defences and destroyed cities such as Sendai when a tsunami hit the coast of Japan last year.

PLAN FOR A TSUNAMI-RESISTANT CITY

Sendai is considering refashioning its coastal area. A raised sea wall would block typical tsunamis and an elevated coastal road would protect against giant ones. A new law mandating zoning restrictions aims to lower the number of fatalities.



who studies Earth and environment interactions at Tohoku Gakuin University in Sendai, examined the trees and found that those that caused the most damage still had their roots. “They didn’t break — they came right out of the ground,” says Miyagi. Trees with roots reaching deeper than 3 metres were usually able to withstand the force of the tsunami, however. “Our recommendation is to build up the land under the trees so they hold,” says Miyagi. This raised land would allow roots to grow longer and would create an extra barrier to protect inland areas.

The trees that stood saved lives in other ways. Some people who missed the call to evacuate were able to climb to safety and wait for rescue. As Imamura drives around Sendai, he points to trees where people rode out the tsunami. He also shows where panicked people took refuge on bridges, schools and river embankments. Many of the official evacuation zones were swept away because they were too low, so the level at which people can be considered safe clearly needs to be revised.

But that is the greatest challenge of the reconstruction, says Hirano. “Even with the general agreement that people should build on higher ground, it’s difficult to agree on where to build,” he says.

SAFETY ZONE

According to the new law mandating tsunami-safe cities, local governments in coastal areas nationwide must simulate a massive tsunami’s effect on the region and then develop zoning policies around the results. Areas where water is likely to reach depths of more than 4 metres are the most dangerous; no residences or hospitals would be allowed to be built in such red zones because people would have difficulty evacuating, especially at night. Offices or factories could be sited there, because workers could easily evacuate. Yellow zones, where water could reach levels of between 2 and 4 metres, are suitable for residences only if they are built on stilts or made with reinforced concrete (see ‘Plan for a tsunami-resistant city’).

Imamura helped Sendai officials to plan the city’s zoning by carrying out some 200 simulations that varied parameters such as the height of the coastal sea wall and the position of a road that serves as an embankment. In Imamura’s most cost-effective scenario, the town’s 6-metre-tall segments of sea wall would be replaced by one long wall 7.2 metres high, which would act as the first line of defence. Behind that, the city would restore a coastal tsunami-control forest 200–400 metres wide and 20–30 kilometres long. Farther inland, the coastal road would be raised from the current 2 metres to 6 metres above sea level.

The region between the coast and the road would be declared a red zone and dotted with artificial hills to serve as evacuation sites for those working there. Other evacuation sites would be located farther inland. Imamura says that this scenario would reduce the inundated area by 60% compared with last year and could, if people were properly trained, avoid deaths.

Although the proposal impressed local officials, many citizens were not happy that some 1,214 hectares and about 2,000 homes would be declared unsafe for habitation. Some have threatened to sue. But Fumio Yamada, head of Sendai’s reconstruction division, agrees with the government’s zoning law and says that he will enforce it. After tsunamis ravaged the town in 1896 and 1933, survivors moved up to the hills but later generations returned. “If you just warn people, if you don’t have it in law, people will come back,” says Yamada.

Yet Iguchi, Iwanuma’s mayor, says he will not kick people off of their land. “I don’t want them to live there. But people have their rights and there have been lawsuits,” the mayor says.

Some researchers worry that the simulations used in zoning decisions are being pushed too far. Satoru Masuda, a specialist in risk communication and disaster-prevention city planning at Tohoku University, says that zoning officials have not given sufficient consideration to the uncertainties of the simulations.

And the debate looks set to spread, says Imamura, as local governments in all coastal

areas conduct the mandatory simulation-based zoning. He sits on a committee expected to report back by next month with a reassessment of the tsunami potential of an earthquake in the Nankai Trough south of Tokyo. When all the maps are adjusted, millions of people will probably fall into red zones and could be told to move, he says. “Even after the disaster, some people in Tohoku are resisting,” says Imamura. “Of course in places where there hasn’t been a tsunami there will be debate. It will take time.”

Imamura hopes that the new institute set to open at Tohoku University on 1 April will help. The International Research Institute of Disaster Science (IRIDeS) will receive ¥800 million per year in funding for ten years, and will have roughly 25 teams. These will analyse the performance of disaster-mitigation technologies; develop better ways to support victims; research early-detection systems for mega-quakes and tsunamis; and establish a medical system for coping with disasters and a digital archive of them, says Imamura, who will be vice-director.

IGNORED WARNINGS

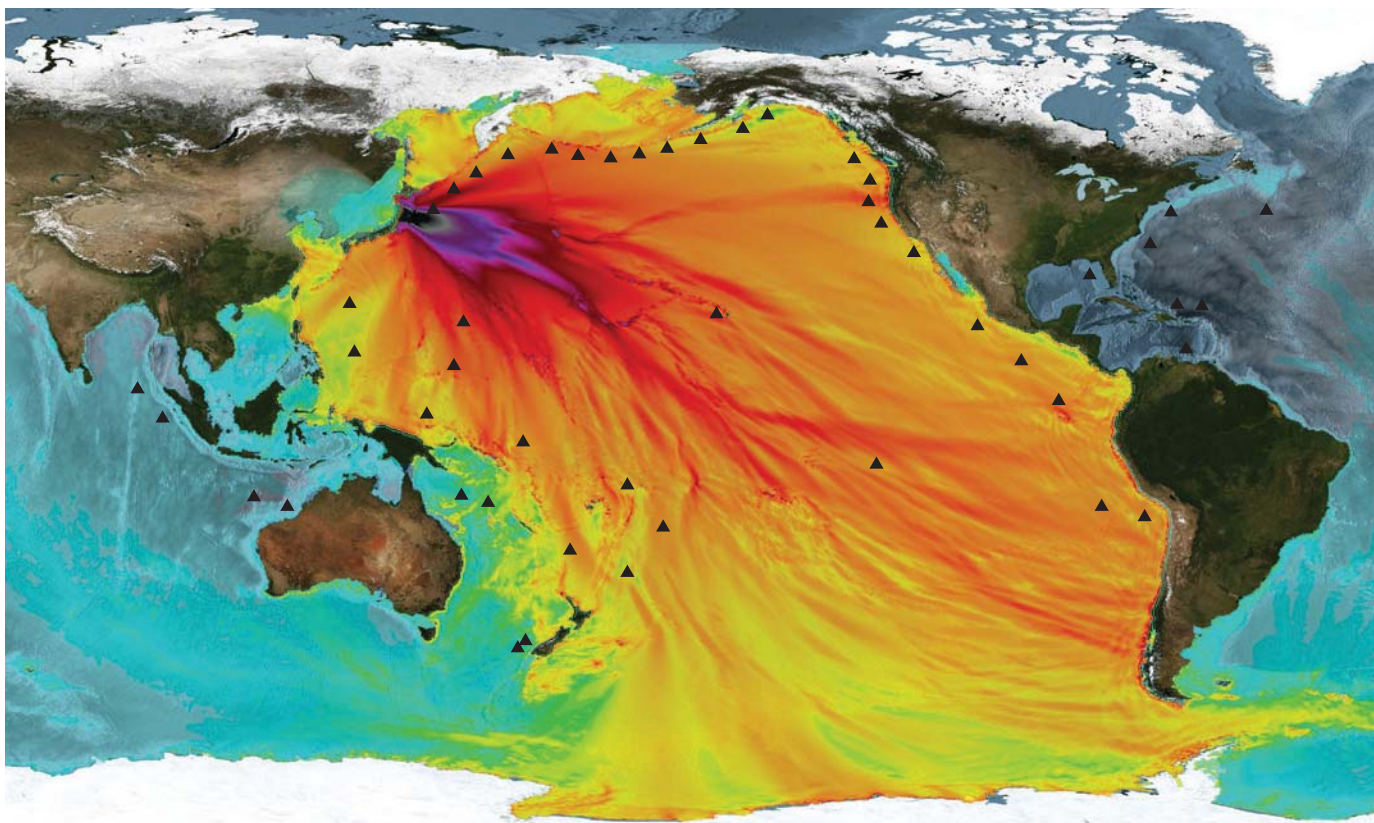
A key focus is how to make people more aware of the danger that tsunamis pose. Tohoku University’s Toshiaki Muramoto, a cognitive psychologist who will take a position in the new institute, plans to look at the way people process information during disasters. He notes that a survey of 870 survivors found that 60% quickly evacuated when they heard the alarms, but the rest waited. Of those, 75% said they “finished what they were doing” first. “There’s a tendency among some to hear a warning and think that they are ok,” says Muramoto. He wants to understand why people usually play down risks and to use that information to devise better educational messages about natural hazards. “That could change things,” he says.

The digital archive at IRIDeS, which will consist of images and records of the tsunami, could serve as a global resource for education and policy planning, says Shosuke Sato, another Tohoku University researcher and future member of the institute. He says the database will be “like Facebook for disasters”.

The real test of the post-disaster research and planning will come when Japan faces its next massive tsunami. In many ways, it will be a memory test — of how well humans, in this age of information, can package and transmit a message that makes future generations conscious of their vulnerability.

It will take an emotional symbol, says a group of some 50 scientists, including Imamura, who propose preserving one of the boats that lodged in a house, an evacuation centre or the twisted warning sign on the beach. “We’re thinking of making spots like this into memorials,” Imamura says, “so that people don’t forget.” ■

David Cyranoski is Nature’s Asia-Pacific correspondent.



NOAA CENTER FOR TSUNAMI RESEARCH

A simulation of the 2011 Tohoku tsunami used data from buoys and sea-floor sensors (triangles) to estimate wave heights. Deeper colours indicate higher waves.

THE NEXT WAVE

What can scientists learn from the Tohoku tragedy to improve tsunami forecasting and save lives?

BY RICHARD MONASTERSKY

As soon as the shaking died down on 11 March last year, Ken-Ichi Sato stumbled back to his office and pressed the alarm button, triggering sirens throughout the city of Kesennuma in northeastern Japan. As the emergency manager of the coastal community, Sato had to alert the 64,000 people there that a tsunami might be coming.

A minute later, that threat became more real when Sato received word from the Japan Meteorological Agency (JMA) that the quake was large — magnitude 7.9 — and located off the coast of Miyagi Prefecture, where Kesennuma is located, in the Tohoku region. Residents there should brace for a 6-metre tsunami, warned the agency, and the neighbouring Iwate and Fukushima prefectures should prepare for waves half that height. Sato immediately issued an evacuation warning over the city's loudspeakers.

But when the tsunami hit around half an hour later, it dwarfed the original JMA estimates. The water surging into Kesennuma reached 9 metres high, and the waves battering other coastal sites topped 20 metres, pouring over the sea walls and barriers that buttress much of the Tohoku coastline. Some 15,000 people died as a result of the tsunami — some of whom had reportedly not fled to higher ground because the projected wave heights had made them think they were safe. In Kesennuma alone, 1,031 people died and hundreds are missing.

Sato thinks that the toll might have been lower had he learned the true size of the tsunami earlier. “We could have raised the

intensity levels of the alerts,” he says. “We could have made sure that people got to a high-enough place.”

A year later, scientists and emergency managers are still struggling to improve their tsunami detection and warning systems before the ocean strikes again. Japan will soon start to install a ¥32.4-billion (US\$402-million) system of ocean-bottom sensors to provide advanced warnings of tsunamis heading towards the coast. And the United States is considering moving some of its deep-ocean warning buoys off the Pacific Northwest coastline closer to the Cascadia subduction zone, where a mammoth quake is expected, perhaps within the next few decades.

The efforts are an extension of advances made since 2004, when a tsunami caused by an earthquake off the coast of Sumatra killed more than 230,000 people. The disaster raised awareness of tsunamis and prompted nations to pump money into research and equipment. As a result, emergency managers can now effectively forecast how tsunamis will cross ocean basins and hit coastlines thousands of kilometres from a quake's source.

The next, more difficult, goal is to improve warnings for close-in regions, which may only have minutes to react. “Historically, maybe 95% of tsunami deaths are from local or regional tsunamis,” says Laura Kong, director of the International Tsunami Information Center in Honolulu, Hawaii. “How do the United States and the global community



A YEAR AFTER THE TSUNAMI

For more content, go to:
www.nature.com/japanquake

address that with the resources we have?”

Japan intends to meet that challenge with its new sensor network, which is designed to keep tabs on the eastern coastline (see ‘Safety net’). Operated by the National Research Institute for Earth Science and Disaster Prevention (NIED) in Tsukuba, the system will consist of 154 sea-floor observatories, each of which contains a seismometer and a water-pressure gauge that can sense the passage of a tsunami, according to Toshihiko Kanazawa, head of the team developing the network. Fibre-optic cables will connect the units in six long loops that each reach the coast at two widely separated locations. According to the NIED, that design will keep the system online even if tsunamis damage a land station or destroy a cable, as happened last March to some sea-floor sensors stationed off the Tohoku coast. The plan is to finish the new network by the end of March 2015.

TRIED AND TRUE

A large cabled network is already in place in the Nankai trough area south of Tokyo, where a large earthquake is expected in the coming decades (see *Nature* 476, 391–392; 2011). The system is “field-proven”, says Kanazawa. “Its simplicity is suitable for tsunami warning use.”

The NIED sensors will sit between the coastline and the earthquake source — the offshore trench where the Pacific plate dives beneath the plate carrying northern Japan. When the Pacific plate jerks forward, the edge of the overlying plate springs up and displaces a huge volume of water, triggering a tsunami. The waves race through the deep water of the open ocean at speeds of roughly 700 kilometres per hour. They alter the sea level by only a metre or two at most as they travel far from the source. But when they hit shallow water, the waves slow down to less than one-twentieth of their former speed and rear up, creating giant surges that sweep ashore. When the NIED network is in place, it will detect the pressure change caused by the tsunami as it travels from the deep ocean over the continental shelf, providing between 5 and 20 minutes of warning for people on the shore.

As a complement to that system, the JMA plans to install three sea-floor sensors on the opposite side of the subduction zone, to catch tsunamis as they speed through the open ocean. Instead of transmitting data through a cable, these sensors will send acoustic signals to nearby buoys that then relay the information up to geostationary satellites. The buoys can be installed faster than the cabled network and will be put in place this year, says Kanazawa.

They will be part of an existing network

of buoys known as DART, for Deep-Ocean Assessment and Reporting of Tsunamis. The United States has 40 such buoys stationed around the Pacific and Atlantic, and other nations have purchased 14 buoys, which are positioned at sites in the Pacific and Indian oceans, with almost all of the data shared internationally.

The impetus to develop the DART system came in part from an expensive false alarm, recalls Eddie Bernard, who designed the system and retired as head of the US National Oceanic and Atmospheric Administration’s Pacific Marine Environmental Laboratory (PMEL) in Seattle, Washington, in 2010. When a magnitude-8.0 quake struck the Aleutian Islands in 1986, Hawaii evacuated its beaches and other low-lying coastal areas at a cost of some \$40 million, including lost revenue. But when the tsunami washed ashore, it was only about 15 centimetres high. The civil-defence leader called up Bernard, who was head of a lab doing tsunami research, and said: “Why can’t you do better?”

The programme got a boost when a 1992 earthquake off the coast of northern California raised concern that the Cascadian subduction zone might let loose with a giant earthquake and trigger a devastating tsunami. So in 1997, Congress provided funds for a tsunami mitigation programme, and Bernard finished his long-term project to develop deep-ocean tsunami sensors.

He and others originally thought of DART buoys as sentinels for the entire ocean, watching for tsunamis from distant earthquakes — the type that usually threaten Hawaii. The buoys had accordingly been stationed far out to sea, both to catch the largest number of tsunamis and because

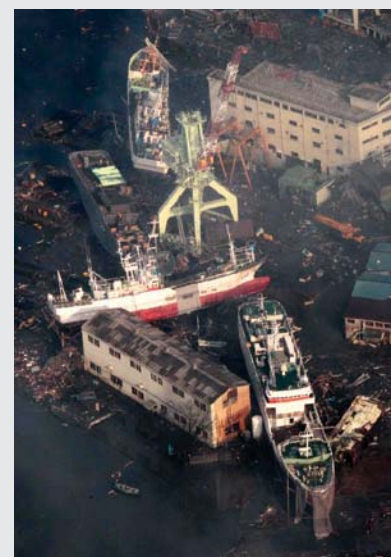
researchers worried that if the sensors were close to the source of

an earthquake, the seismic vibrations would drown out any tsunami signal.

But the Tohoku quake changed that thinking. At a meeting in January at the PMEL, Japanese and US scientists discussed ways to filter out the seismic vibrations from the sensor data, which would mean that the sensors could be deployed much closer to faults, says Vasily Titov, a tsunami modeller at the PMEL. “We can put the sensors in a place 5 minutes away from the sources,” he says. Once the front of the tsunami reaches a DART sensor, Titov says, it takes another 5–10 minutes for half of the wave to pass by, thereby revealing the height of the tsunami.

The United States is now hoping to move some of its DART buoys nearer to earthquake sources — along the Cascadia subduction zone and in many other regions, says Titov.

“WITHIN HALF AN HOUR, YOU CAN GET A VERY HIGH-QUALITY FORECAST SHOWING WHICH AREAS ARE GOING TO BE INUNDATED.”



Kesennuma was hit hard by the tsunami.

WARNING SIGNS

In the minutes after the Tohoku earthquake on 11 March last year, authorities in the city of Kesennuma in Miyagi Prefecture used the available data to issue a series of tsunami warnings. They quickly called for evacuations, but in the end, more than a thousand people in the city died.



- **14:46** Earthquake hits off the Tohoku coast.
- **14:48** Ken-ichi Sato, emergency manager of Kesennuma, sounds alarm of a possible tsunami.
- **14:49** Japan Meteorological Agency (JMA) measures quake as magnitude 7.9 and issues a warning for tsunamis with heights of 6 metres in Miyagi Prefecture and 3 metres in the Iwate and Fukushima prefectures.
- **14:52** Sato issues tsunami warning, telling people to evacuate.
- **14:58** Sea-floor pressure sensors detect a tsunami roughly 50 kilometres off the coast of Iwate.
- **15:10** A sensor 20 kilometres off the Iwate coast detects a tsunami.
- **15:12** Tide gauge on Iwate coastline records major inundation.
- **15:14** JMA updates warning for 10-metre wave for Miyagi and 6 metres for Iwate and Fukushima.
- **15:21** Water levels start to rise in Kesennuma, after an initial lowering caused by the tsunami.
- **15:33** Tsunami in Kesennuma reaches height of 9 metres.

47 MINS



Warning centres will then combine the DART data with spatial models of coastlines to predict the severity of the flooding more quickly. "Within half an hour, you can get a very high-quality forecast showing which areas are going to be inundated," he says. And that would help emergency managers to decide which areas should be evacuated, and when to urge people to move to higher ground.

In the regions closest to an earthquake, however, many people would die if they waited for those results. The first waves from a Cascadia tsunami can hit in 15–20 minutes, and the problem is even worse in Japan and the Aleutian Islands, where some regions have only a few minutes of lead time. Emergency managers are therefore developing a tiered approach, in which they issue quick warnings that are updated as measurements come in from the sea-floor sensors.

UPS AND DOWNS

The Tohoku earthquake shows how those data can help — and hurt. At an international meeting last month in Sendai, Japan, Osamu Kamigaichi from the JMA described some of the problems that his agency ran into during the disaster. When the quake struck at 2:46 p.m. local time, the agency quickly determined its size and location from records of short-period seismic waves, the first data to become available. The JMA then used pre-computed tsunami simulations for the estimated quake to forecast the height of the waves. The warning with those details went out within 3 minutes.

This method works well for quakes smaller than magnitude 8, but it can't gauge the size of larger shocks. The JMA didn't consider that a problem, however, because the estimated size of the Tohoku event was 7.9, about the size of the largest earthquake expected there. But the quake turned out to have a magnitude of 9.0, more than ten times stronger.

The first hints that something was amiss came at 2:58 p.m., 9 minutes after the first warning, when a cabled pressure sensor picked up an unexpectedly large change in sea level off Iwate prefecture (see 'Warning signs'). But the agency did not have a fully developed method for using data from that sensor to update the tsunami warnings.

At 3:10 p.m., a Global Positioning System sensor off the Iwate coast also detected a large tsunami. The JMA used that reading to estimate how much the tsunami would grow when it hit the shallow coastal waters. At 3:14 p.m., an upgraded warning went out predicting 10-metre tsunamis in Miyagi and 6 metres in Iwate and Fukushima. But by then, the first waves had already slammed into the coast.

The JMA may also have caused confusion when it released the initial wave amplitude from a coastal tide gauge, which was 20 centimetres. The earliest waves in a tsunami are not always the largest, and that early statement could have caused people to delay or even halt

SAFETY NET

To detect tsunamis, Japan plans to deploy 154 observing posts linked by sea-floor cables by 2016. This year, it will install three buoys that will relay information from deep-sea tsunami sensors.



their evacuations, says Kamigaichi.

The agency plans to adopt a new tsunami warning procedure by the end of the year. It is developing an analysis tool to tell whether the quick method is likely to underestimate the size of the quake. The tool uses recordings of the strongest vibrations from a broad area and the early measurements of long-period vibrations. If that additional information indicates that the estimated size is accurate, then the JMA will issue a warning that includes the size of the expected tsunami. But if the quake estimate seems inaccurate, the agency will issue a warning for the worst-case scenario — based on historical data in the area — and will use only qualitative descriptions such as 'huge' or 'large' to describe the anticipated tsunami.

The agency will then update the warning about 15 minutes later, once the data have arrived from offshore tsunami sensors.

SAVING LIVES

Modellers in the United States and Japan are confident that their work will eventually pay off. Kenji Hirata, a senior researcher at the JMA, says that he needs several years to finish developing the algorithms for assimilating data from offshore tsunami sensors into forecasting tools for close-by earthquakes.

The work may be particularly useful for regions outside the immediate vicinity of the quake, where people may have an hour or more before the tsunami hits but could still face mammoth waves. In those cases,

the high-tech data could help emergency managers to decide whether an evacuation is justified.

Even for areas where the first wave will hit quickly, direct measurements could help because they can give an idea of whether subsequent waves will be bigger or smaller. And they can provide advance warning of when a relatively small earthquake has triggered an underwater landslide that might then spawn a large tsunami. That sequence happened after a magnitude-7.1 quake in Papua New Guinea in 1998, and it killed more than 2,000 people, many of whom didn't evacuate because the earthquake was not particularly large. In such cases, offshore tsunami sensors might save people because they can pick up events that seismic readings miss.

Researchers caution, however, that no amount of expensive hardware can replace basic education about tsunamis. In many cases, communities simply won't have time to wait for estimates of the tsunami's size. "If you live in a coastal area, you have to be your own warning centre," says Costas Synolakis, a tsunami researcher at the University of Southern California in Los Angeles. "If the earthquake lasts for more than 30 seconds, it means it's a big earthquake and local, and you have to evacuate. If it lasts for over 2 minutes," he says, "it means run for your life. This is a giant." ■

Richard Monastersky is a features editor for *Nature* based in Washington DC. Additional reporting from **David Cyranoski** in Japan.

COMMENT

SEISMOLOGY Lessons learned from the megaquakes of the past decade **p.149**

NUCLEAR ENERGY Economics should dictate how we respond to Fukushima **p.151**

PHOTOGRAPHY The woman who caught physics on film **p.156**



NEUROSCIENCE Pianist Vijay Iyer riffs about jazz and cognition **p.157**

MIYAKO CITY OFFICE/REUTERS



The 2011 Japan tsunami breached 10-metre-high walls in Iwate Prefecture, where warnings predicted a wave only 3 metres high.

Putting seismic research to most effective use

Today's tools and geophysical knowledge could be utilized more effectively for earthquake hazard mitigation, says **Hiroo Kanamori**.

A year has passed since a devastating tsunami inundated large areas of the northeastern coast of Japan, with tragic consequences. Given that it remains too difficult to make accurate short-term predictions of earthquakes, at least for now, the question is what we can do to improve earthquake damage mitigation.

The 2011 magnitude-9 Tohoku-Oki earthquake clearly demonstrated that seismologists can rapidly and quantitatively determine what has happened during a

quake, thus reducing its impact and the number of lives lost. Yet there is much room for improvement. Real-time data are not being fully shared between nations, and the best monitoring tools have not been fully implemented. More should be done to encourage the practical use of seismic

research. At present, there is too large a gap between academic work and the reality of hazard mitigation.

Thanks to the extensive global seismic network established over the past three decades, the world's seismic activity is constantly monitored in real time by the National Earthquake Information Center (NEIC) of the US Geological Survey. Some 20 minutes to an hour after the onset of the Tohoku-Oki earthquake, the NEIC had located the epicentre and determined that the ►



A YEAR AFTER THE TSUNAMI
For more content, go to:
www.nature.com/japanquake

► magnitude was about 9, with a faulting mechanism typical of subduction-zone great earthquakes¹. This information was crucial for alerting the emergency services. Such a rapid determination was not possible until recently; in 2004, it took hours or days for similar warnings to be issued¹.

Even more rapid determinations, within just a few minutes of an earthquake, can make a great difference to local warnings. Within 3 minutes of the Tohoku-Oki quake, the Japan Meteorological Agency (JMA) system came up with an initial estimate of magnitude 7.9 for the quake. On this basis the agency issued a warning of 6-metre tsunamis along the coast of Miyagi Prefecture, and 3-metre tsunamis along the coast of neighbouring Iwate and Fukushima prefectures. Exactly how many lives were saved by this rapid warning is hard to determine; perhaps many thousands². Unfortunately, some residents in Iwate Prefecture, where tsunami walls as high as 10 metres had been built, assumed that they would be safe from tsunamis less than 3 metres tall and did not take immediate action. In the end, the tsunami overtopped many of these walls.

ROOM FOR IMPROVEMENT

Forecasters could improve warnings simply by using existing technology better. Crucially, a rapid tsunami-warning system needs to have access to high-quality data at short distances, and a method that can determine both the magnitude of the earthquake and the mechanism of faulting. The latter requires the use of both short-period seismic wave information and, for large earthquakes, longer-period wave information. The JMA has high-quality data, but has not yet implemented the best methodology: its current system uses only

short-period information, which led to the initial underestimate of the magnitude of the Tohoku-Oki quake. It has been testing a better system offline since 2008, but has not yet put it into action. The NEIC, by contrast, has a good methodology, but does not have sufficient access to real-time local data. Although Japan is covered with hundreds of broadband seismographs that are suitable for warning of tsunamis, the data from only a handful are made available to the NEIC in real time. The number of stations worldwide sending real-time data to the NEIC has gone up from 350 in 2004 to 1,183 in 2011

(ref. 1). Hopefully this trend will continue.

Ideally, both the NEIC and JMA should improve their systems in those aspects in which they are lacking, so that we can have backup systems in place for verifying the impact of ongoing earthquakes. This is especially important when the operation centres in the affected country are incapacitated by a natural disaster.

In addition to seismic data, high-rate Global Positioning System (GPS) data that can reveal the scale of large ground motions are now available. Several research groups are developing new methods for using real-time GPS data in rapid warnings. These should enable an estimation of fault size in addition to earthquake magnitude, and could therefore produce better tsunami warnings. In Japan, for example, this methodology could produce accurate warnings within just a few minutes.

Real-time international exchange of both seismic and GPS data is therefore vital for

more rapid, reliable and robust warnings.

Japan also has a nationwide earthquake early-warning system, which can give 1–30 seconds of warning (compared with a warning time of 5–30 minutes for a tsunami) by detecting the first signs of shaking. This was successfully deployed during the Tohoku-Oki earthquake, although the extremely complex rupture process, along with the many aftershocks, confused the system and forced it to be shut down temporarily.

The most effective application of this type of early warning is when it is paired with engineered systems that will automatically shut down in the event of strong shaking, as with the Japanese bullet trains. At the time of the Tohoku-Oki earthquake, about 24 trains were running in the affected area, some at speeds in excess of 200 kilometres an hour. Emergency braking started within a few seconds, bringing all the trains to a halt without a single derailment, significantly damaged train, serious injury or death. Similar automated systems are now in use by some private companies.

It is essential that seismologists work closely with engineers to optimize these and similar systems. The merit of many of the academic theories and methods produced by seismologists may not be obvious to industry, and there are few financial incentives to make them more practical. By contrast, in the materials and biomedical sciences, good products developed in academia are much more likely to be picked up by industry because they can be developed into profitable products. It is important for seismologists and geophysicists to take a more active role in promoting interaction and collaboration between disciplines and with industry.

It is equally important to cultivate good human resources. One lesson learned from the Tohoku-Oki earthquake is that nature occasionally acts in ways that contradict our assumptions (see page 149). It is important to train students of science to have flexible and creative minds that can challenge the conventional view (about where large earthquakes can happen, for example) before a disaster forces a rethink. As many scientific projects get bigger, young scientists tend to be overwhelmed by the highly technical issues. We need to encourage them to take a step back and to see the big picture, rather than becoming overly specialized at a young age or locked into conventional ways of thinking. ■

Hiroo Kanamori is a seismologist at the California Institute of Technology in Pasadena, California 91125, USA.
e-mail: hiroo@gps.caltech.edu

1. Hayes, G. P. et al. *Seismol. Res. Lett.* **82**, 481–493 (2011).
2. Heki, K. *Science* **332**, 1390–1391 (2011).



The Japan Meteorological Agency issues tsunami warnings within minutes of an earthquake.



The 2011 Tohoku-Oki earthquake that shook Japan was the latest in a string of surprising megaquakes.

Why giant earthquakes keep catching us out

A spate of huge earthquakes in the past seven years has provided humbling lessons, says **Thorne Lay**.

If you are under the impression that there have been an unusual number of horrific earthquakes of late, often contradicting scientific expectations, you are correct. The rate of occurrence of 'great' earthquakes — those of magnitude 8 or larger — since December 2004 is about 2.5 times the average rate over the preceding century. Five quakes of magnitude 8.5 or greater have struck since the magnitude-9.2 Sumatra–Andaman earthquake on 26 December 2004, which produced a devastating tsunami that took more than 230,000 lives around the Indian Ocean. One year ago this month, the magnitude-9 Tohoku–Oki earthquake and tsunami in Japan caused massive devastation.

This activity is not unprecedented in the long term: from 1950 to 1965, Earth was wracked by seven earthquakes of magnitude 8.5 or greater, including the largest yet recorded — magnitude 9.5 in Chile in 1960. Where the recent batch of quakes differs from historical ones is in seismologists' ability to quantify the rupture processes. Modern instruments for recording ground motions, and the analytical methods for extracting information, were developed mostly between 1970 and 2000 — a quiet interval with few great earthquakes and none over 8.5. In effect, the geophysical community became ready to study great earthquakes just in time for the recent batch.

Almost all the recent events have violated some theories of where and when great earthquakes can occur and what their consequences can be. This is perhaps unsurprising, given the short time over which detailed data on such events have been available.

The recent great earthquakes have all hit near subduction zones: regions in which oceanic plates are diving under other plates. Here, friction causes sticking and slipping of the rocks that results in intermittent earthquakes. Geophysicists thought that in certain regions of these plate boundaries, warm rock or slippery sediments should prevent the build up of friction enough to avoid a large quake. We thought that regions that had ruptured recently wouldn't rupture again for years, and that segments where a long time had passed since the last big event were most likely to slip in the near future. We underestimated the extent to which one earthquake can trigger another. Our analysis procedures had been developed to handle faulting that lasts seconds rather than the minutes seen in some recent quakes.

Researchers have long tried to make rational decisions about which fault zones

should provoke the most concern, to focus limited resources for earthquake preparedness. But having only a century's worth of detailed earthquake history, we have sometimes been lured into ignoring areas that were harbouring potential for giant slip. It is essential that geophysicists learn as much as possible, as quickly as possible, from the recent events. Population growth means that more people are now, and will be, exposed to earthquake hazards than during 1950–65.

TRIGGER POINTS

A common question is whether the recent great earthquakes are directly related to each other. The record over the past 110 years suggests that such spates of activity may be a statistical fluctuation. Indeed, there is no corresponding spate of increased occurrence of smaller events such as the devastating magnitude-7 Haiti earthquake of 12 January 2010. But great earthquakes can interact on a regional scale. The 2004 Sumatra–Andaman earthquake, for example, is thought to have triggered further great quakes along the same plate boundary in 2005 and 2007 (see 'Surprise scenarios'). A stretch of the Sumatra plate boundary between these quakes has not ruptured since 1797 and seems to be stuck, accumulating strain — an area of concern for a future great earthquake.

Whether a large earthquake can have a wider reach, triggering activity on the other side of the planet that leads to another great earthquake years later, is controversial. So far, seismic waves from a major quake have been seen to directly trigger only small events at a great distance. However, it is conceivable that dynamically triggered events in a far-off region could induce a great earthquake there sooner than would otherwise have been the case.

We still have much to learn about earthquake triggering. On 15 November 2006, for example, the plate boundary between the sinking Pacific plate and the overriding plate along the Kuril Islands (which run between the Kamchatka Peninsula and Japan) ruptured in a magnitude-8.4 earthquake. This triggered small earthquakes 100 kilometres offshore in the Pacific plate — a common occurrence. But two months later, a magnitude-8.1 earthquake shook the same region. There is no straightforward theory that predicts how often or when such a great aftershock might occur. An even more exotic case of triggering took place on 29 September 2009. A magnitude-8 event within the Pacific plate under the Tonga trench occurred first, triggering a second magnitude-8 rupture of the plate boundary 50 kilometres to the south, even while the first rupture was still growing. This pair produced a large tsunami that caused extensive damage in Samoa, American Samoa and Tonga.

Other earthquakes have brought sobering



A YEAR AFTER THE TSUNAMI

For more content, go to:
www.nature.com/japanquake

lessons about our ability to issue early warnings. Some researchers have argued that quakes grow from an initial rupture that has some kind of signature indicating how large the final rupture will be. Recent earthquakes have undermined that hope. Great events in Peru in 2007 (magnitude 8.0) and 2001 (magnitude 8.4) both seem to have started with magnitude-7.7 ruptures that, after a pause, triggered adjacent regions to fail with larger energy release. The 2011 Tohoku-Oki earthquake (magnitude 9.0) seems to have begun as a tiny magnitude-4.9 event. Early-warning efforts are challenged by these erratic growth patterns.

RUPTURE MECHANICS

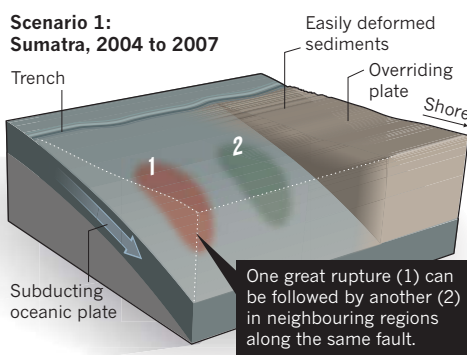
The first surprising megaquake of the past decade, the 2004 Sumatra–Andaman event, ruptured along an unexpected 1,300 kilometres of the plate boundary, extending into an area where seismologists had thought that the plates would shear horizontally rather than subduct. It was extraordinarily long lasting: fault sliding continued for more than 7.5 minutes. Most procedures for determining earthquake slip had been developed using smaller events, lasting perhaps 30 seconds, for which remote seismographs record two distinct signals: primary (P) waves followed several minutes later by secondary (S) waves. In the Sumatra event, P waves were still coming in as the S waves from the start of the rupture arrived, muddling the two. Yet it took only a few weeks for research groups to modify their algorithms. Such updated programs are now widely used, allowing researchers and government agencies to analyse the faulting of even the largest earthquake within minutes to a few hours of an event.

The magnitude-8.1 earthquake on the plate boundary in the Solomon Islands on 1 April 2007 (2 April local time) ruptured an unexpected location, at which an oceanic rift system — a ridge where new ocean crust is being formed — intersected the subduction zone. Convention held that rift-zone rocks, being hot and deformable, would not accumulate enough frictional strain for a large quake. However, analysis of seismic waves showed that the two sinking plates slipped under the overriding Pacific plate in different directions; the rupture spread right across the ‘triple-junction’ at which the three plates came together. Re-evaluation of risk assessments for other areas where ridges are subducting, from Peru to Tonga to Vanuatu, is under way.

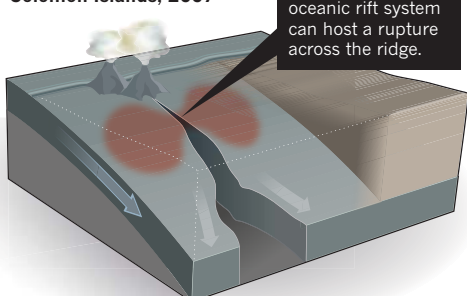
The magnitude-8.8 earthquake on 27 February 2010 in Chile struck a section of the plate boundary that last ruptured in 1835 (Charles Darwin experienced the shaking). Basing our ideas of where large

SURPRISE SCENARIOS

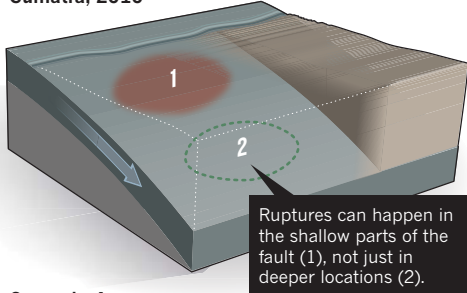
Scenario 1: Sumatra, 2004 to 2007



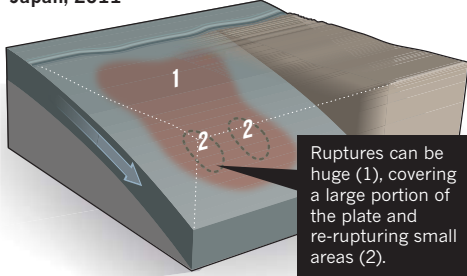
Scenario 2: Solomon Islands, 2007



Scenario 3: Sumatra, 2010



Scenario 4: Japan, 2011



earthquakes will occur on historical events has fostered the concept that plate boundaries are strongly segmented, with isolated regions repeatedly experiencing comparably sized events. This is not always the case. Although the 2010 earthquake initiated in the central region of the 1835 rupture zone, where large slip was expected, the main sliding was located to the north, where the plate boundary had ruptured as recently as 1928.

A magnitude-7.8 earthquake off the coast of Sumatra on 25 October 2010 was surprising in that it ruptured the shallowest portion of the subduction zone, creating a huge

tsunami. The low-strength, easily deformed sediments on the plate boundary at shallow depth in the fault should, it was thought, prevent the accumulation of elastic strain and thus stop large slip. They did not. We must now consider this type of event in risk assessments, because it shows that large tsunami-generating events could happen nearer to the ocean trench than previously thought.

Similarly, the 11 March 2011 Tohoku-Oki event occurred on a plate boundary that historically had experienced much smaller events. The shallow portion of the fault was not expected to accumulate a large amount of elastic strain. Surprisingly, the event had huge slip of 40–80 metres at shallow depth, and also drove slip in adjacent fault segments that had ruptured as recently as 2005, increasing the total size of the event and producing a larger tsunami than had been prepared for.

BITTER LESSONS

Unfortunately, we often advance our understanding of earthquakes through bitter experience. But progress is being made.

We are improving our ability to anticipate risk in specific regions. Geodetic observations of strain accumulation were already mapping out regions with potential for large earthquakes in Sumatra, Chile and Japan before recent great earthquakes struck there. Sea-floor measurements off Japan were on the verge of defining the potential for the large, shallow fault displacements that occurred in 2011. Given more time to observe where strain was accumulating, the size of the earthquake and tsunami might not have been a surprise. Even as things were, rapid analysis of data from regional seismic networks in Japan enabled early warnings; warnings that can be improved (see page 144).

Part of the solution is a wider spread of instrumentation. The dense coverage of seismometers and Global Positioning System instruments in Japan has helped immensely in monitoring activity and in understanding the recent great event. Other regions need to be better instrumented, especially offshore. A major effort is under way to improve our understanding of great-earthquake risk off the coasts of Oregon and Washington in the United States, where a magnitude-9 earthquake last struck in 1700.

Although individual great earthquakes are still sure to surprise us, the most important shift in our thinking has already occurred: we must allow for the possibility of larger earthquakes in regions where we thought that potential did not exist. That is one difficult lesson that we can consider learned. ■

Thorne Lay is in the Department of Earth and Planetary Sciences, University of California Santa Cruz, California, USA.
e-mail: tlay@ucsc.edu

➔ **NATURE.COM**
To hear a podcast by Thorne Lay on this article, see:
go.nature.com/b2ajob



The Olkiluoto 3 nuclear plant in Finland is four years behind schedule and some €2 billion over budget.

The nuclear landscape

The accident at Fukushima has convinced many nations to phase out nuclear power. Economics will be the deciding force, says **Peter Bradford**.

A year after the nuclear meltdowns at the Fukushima Daiichi plant in Japan, tens of thousands of people from the area still cannot return to their homes. Many will not be able to do so during their lifetimes. Only two of Japan's 54 reactors are operating, and there is uncertainty over whether the others will return to service. Once-robust Japanese plans for new nuclear construction have been abandoned, with the government's energy white paper in October 2011 calling for a reduction in reliance on nuclear power.

The disaster has triggered other countries to rethink their nuclear-energy policies. Germany, Belgium and Switzerland have announced plans to close all of their existing reactors (although phase-out plans have been reversed before). They will not build new reactors. Neither will Italy, which closed its reactors after the 1986 Chernobyl accident in what is now Ukraine. In June 2011, some 94% of the Italian electorate voted against any resumption. The Czech Republic has cancelled all but two of its planned plants. Poland has deferred reactor construction for several years. Even France, long a nuclear-power stalwart, may close some reactors early depending on the outcome of elections later this year.

Other nations have made less drastic changes, such as bringing in enhanced safety measures or 'stress tests' for reactors. In countries including the United Kingdom, the United States, China, Russia, India and South Korea, plant closures and cancellations attributed to Fukushima are less likely, but planned expansions will struggle to stay on course.

Before Fukushima, the number of new nuclear plants built in the past 20 years roughly equalled the number retired. The latest re-evaluations suggest at least a short-term downturn for nuclear power's share of global energy production.

Yet the most implacable enemy of nuclear power in the past 30 years has been the risk not to public health, but to investors' wallets. No new nuclear-power project has ever bid successfully in a competitive energy market anywhere in the world.

Fukushima is not the only recent challenge to nuclear power: long-term natural-gas prices and price forecasts have more than

halved in the United States, and globally, electricity demand has dropped on average owing to the economic slowdown. But these events have merely worsened the odds for new nuclear reactors: private capital was no more available before these events than it is now. Accidents at nuclear plants and economic changes have occurred before, within the living memory of global financial communities. The knowledge that any of these events — or new future threats — might happen has long discouraged investors.

Buffeted by economic forces, new nuclear capacity cannot be expected to contribute significantly to global energy supplies or to climate-policy solutions in the decades ahead. Most of the world's reactors are more than 20 years old, so plant completions will be largely offset by retirements. Wise economic and environmental policy will have to allow nuclear energy whatever place it can earn among other energy sources, and not burden it with unnecessary hopes and fears.

THE SECOND COMING

In the United States, where no new reactor has received a construction permit since 1978, a 'second coming' of nuclear construction has been anticipated with as much ballyhoo (and accuracy) as its religious counterpart. In 2009, 31 applications for new reactor projects had been filed or announced. Three years later, only four projects remain on course for construction by 2021. Only on 9 February this year did the US Nuclear Regulatory Commission issue its first licence to build a new reactor in 34 years.

Many people have accepted a version of history in which US nuclear construction wound down because of excessive regulation, environmental alarmism and public fearfulness after the 1979 accident at Three Mile Island in Pennsylvania. For them, the new reactor licence for a site in Georgia proves that US regulators and the public have regained their resolve. But this view rests on a fallacy. Three Mile Island happened at about the same time that competitive market forces were unleashed on the process of power-plant selection. This is the true reason for the decline in reactor construction: some 50 US reactor orders had already been cancelled before the accident.

Chicago-based firm Exelon, the largest US nuclear operator, predicts that electricity from new nuclear plants will cost 12 cents per kilowatt hour (kWh). In comparison, other low-carbon alternatives — combining natural gas with energy efficiency, expansion of existing nuclear plants and wind — will cost 3–11 cents per kWh. Trying to sell new reactors into US power markets that do not value low carbon, and that are predicted to pay no more than 6 cents per kWh for many years,



A YEAR AFTER THE TSUNAMI

For more content, go to:
www.nature.com/japanquake

is a certain route to bankruptcy.

This economic impracticality has been felt globally since the early 1990s. Although the number of nuclear plants in operation jumped from a handful in 1960 to 424 in 1989, they then levelled off, peaking at 444 reactors in 2002 (see 'Not going up'). After Chernobyl, at least a dozen plants of the same design were taken out of action. Older, smaller plants were decommissioned. But the determining factor was the drop in construction in the United States and western Europe — a result of the cost of building new plants and the refusal of investors to bear the risks of cancellations, cost overruns and the emergence of cheaper alternatives.

At the time of the Fukushima disaster, only four countries (China, Russia, India and South Korea) were building more than two reactors. In these four nations, citizens pay for the new reactors the government chooses to build through direct subsidies or energy price hikes. In a few cases, such as the Olkiluoto 3 reactor under construction in Finland, the company providing the reactor promises a fixed price or looks to its own government for finance. Olkiluoto is being built by Areva, an entity largely owned by the French government, which was gambling that the project would jump-start demand for its newest reactor design. As Olkiluoto is four years behind schedule and more than €2 billion (US\$2.7 billion) over budget, that gamble has fallen flat.

The only nations for which building high-cost nuclear plants has any prospect of becoming competitive are those that combine very limited access to natural gas with high growth (Japan once among them). Even then, there is no reason to pay over the odds for nuclear. Instead, all greenhouse-gas mitigating measures and alternative-energy sources should have to compete with each other in the marketplace.

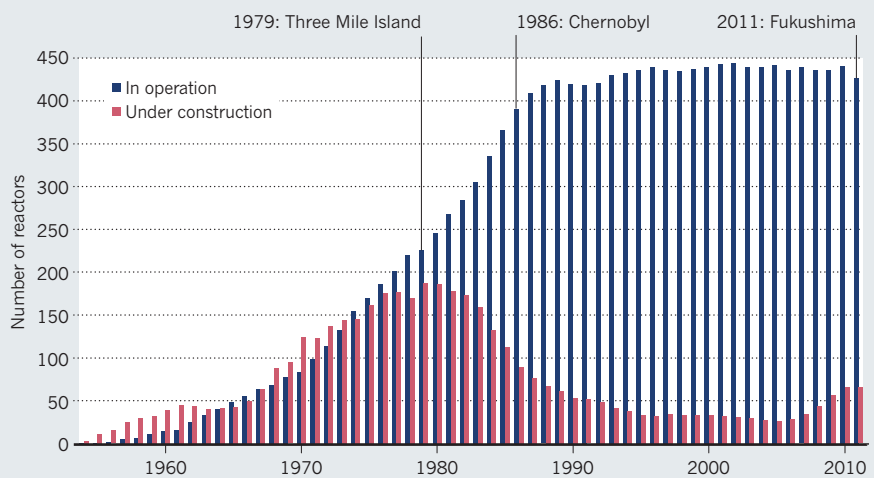
POLITICAL BALANCE

On top of the ongoing dire economic situation, Fukushima has shifted the political equation. The Chernobyl disaster was widely seen as a product of the lax Soviet safety culture, in which an unsafe design was combined with reckless operation. But Japan was considered to be in the technological and regulatory front rank. Champions of nuclear power in the US Congress (and even President Barack Obama on one occasion) had held up Japan as a nuclear success story that US regulators and energy policy-makers should strive to emulate.

After Chernobyl, Italy and Germany

NOT GOING UP

Building of new nuclear reactors declined after the 1979 Three Mile Island accident, as economic realities hit home. The number in operation has levelled out since the late 1980s.



SOURCE: IAEA-PRIS, WSC, 2011

decided to phase out nuclear energy, as Sweden had done after Three Mile Island. All three later reconsidered as concerns over climate change and years without a conspicuous reactor accident shifted the political balance back towards nuclear energy. The unfavourable economics did not change, but the willingness of governments to override economics with subsidies and mandated purchases did.

In the United States, this change has produced a paradox. The right-wing legislators who swept into state and federal office in 2010 have conflicting tendencies towards nuclear energy: they are largely pro-nuclear, pro-market, climate-change sceptics, yet they disparage governmental backing of particular technologies with taxpayers' money. Before Fukushima, the net effect seemed to be towards state and federal measures that were favourable for new nuclear capacity, such as increases in taxpayer-backed loan guarantees or laws that made energy customers responsible for runaway costs or cancelled plants. After Fukushima, those measures quietly died.

In the United Kingdom, too, a conservative government pledged to produce new nuclear power without government subsidies. It is now having to torture the language of new policies to subsidize new reactors without this being recognized as such.

It is impossible to say what will happen next for nuclear energy. The indications from emerging carbon markets are that they will not produce a high enough price to support new nuclear. These markets are not providing enough emissions avoidance either, so the price for carbon could go higher. If Exelon's predictions are correct, and if gas prices remain low, carbon prices would have to more than triple to make new nuclear look economical. At that level,

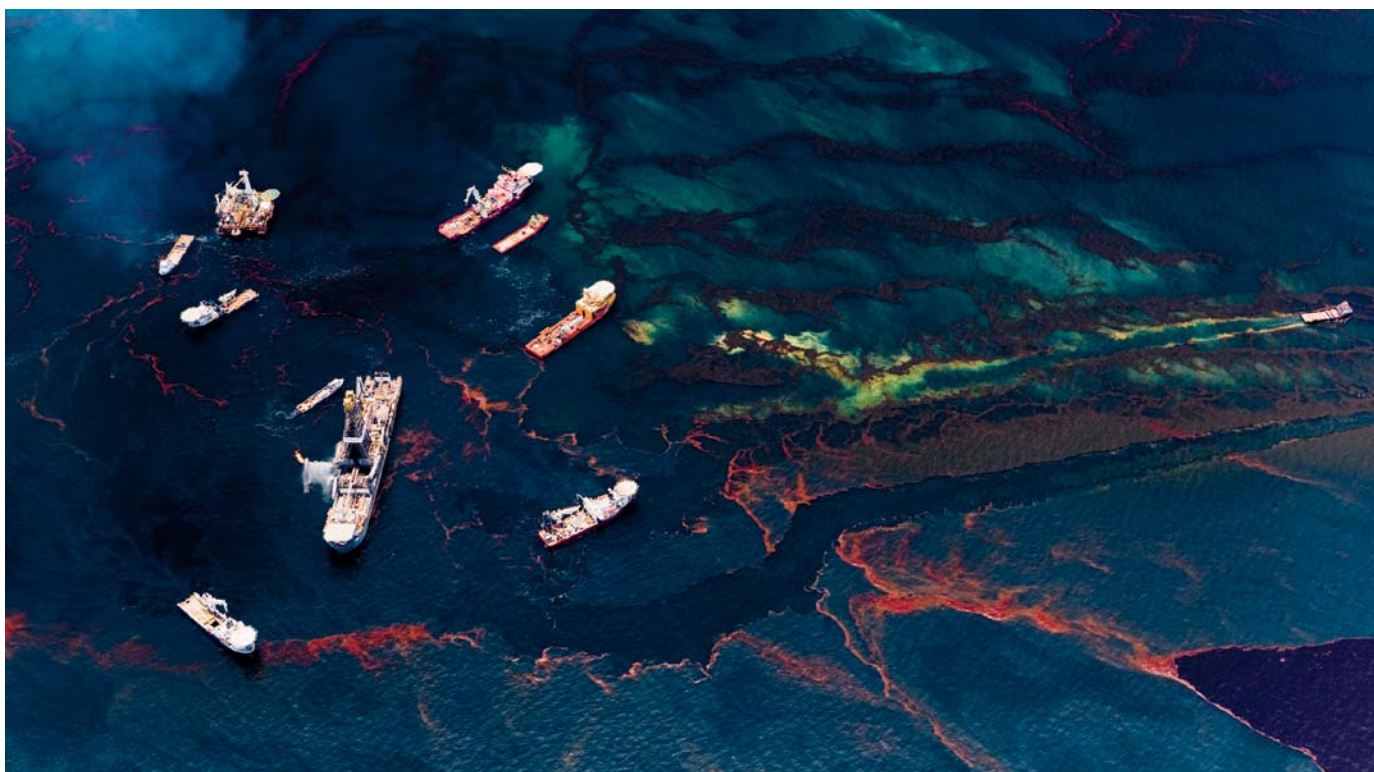
many other alternatives could be built on a large scale more quickly.

The hopes of the US nuclear industry now rest on the new reactor in Georgia — one of the few states with a law assuring that energy customers will pay all reactor costs regardless of price. Energy secretary Steven Chu, visiting the site last month, intoned: "If this project goes forward and is built on-budget, on-time and on-schedule, that would be a very good thing. A lot of other companies will say, 'OK. We now know we can do this and it would be a good investment.'" But companies will say no such thing unless they can also produce energy at a competitive price. Instead, they will join Exelon in saying 'we can do this and go broke'.

Many countries state the unknowable or the implausible with great certainty. Germany has declared that existing nuclear capacity is too dangerous and can be phased out at little cost. The Middle East and the Czech Republic assert that new nuclear is a vital low-carbon energy source. Others, including the United Kingdom, have said that there will soon be new nuclear without subsidies.

It would be ideal if Fukushima could steer us away from these prophecies and towards a sensible assessment of market economics, climate science and nuclear risks. Then nuclear power would serve the public, not the other way around. I don't know how many reactors we would get, but we would get the number that we need. ■

Peter Bradford is an adjunct professor teaching nuclear power and public policy at Vermont Law School, South Royalton, Vermont 05068, USA. He served on the US Nuclear Regulatory Commission in 1977–82. e-mail: perubrad@aol.com



D. BELTRA/GREENPEACE/NATUREPL.COM

Oil on the waters of the Gulf of Mexico could alter gene expression in fish and other marine organisms for years to come.

ENERGY

Plumbing the depths

A chronicle of events preceding the Deepwater Horizon oil spill in the Gulf of Mexico has a thriller-like edge, finds **Amanda Mascarelli**.

Many people have wondered how the Deepwater Horizon oil spill of 2010 happened, given the technological know-how and industry back-up systems in place to prevent such a catastrophe. In *Run To Failure*, journalist Abraham Lustgarten explains how, in his view, the blowout was not only predictable, but inevitable. The book is one of the few that focus on the roots of the disaster — and is refreshingly different from the many looking mainly at the immediate tragedy and its aftermath.

Run to Failure reads like a thriller, complete with whistle-blowers and double agents. On the basis of meticulous investigative reporting during and after the spill, analysis of BP documents and e-mails, and interviews with current and former industry insiders, Lustgarten paints a picture of neglect, hollow proclamations about safety and environmental stewardship, and draconian cost-trimming going back two decades. That history, he believes, helped to pave the way to the disaster.

The Gulf of Mexico has long been seen

as the frontier of oil exploration. In the mid-1990s, most of the oil extracted in the region came from shallow waters. But deep-water wells are much more bountiful. To push into the depths, BP invested tens of billions of dollars in the Gulf, says Lustgarten, and sent in its most technologically advanced drilling rigs and resources. However, to subsidize its investment, Lustgarten alleges, BP made cuts elsewhere, including at its ageing US refineries and on the Alaska North Slope — which had been the source of much of BP's oil production since the 1970s, but where reserves were dwindling.

The oil industry as a whole has a patchy record on safety and the environment, and



Run To Failure: BP and the Making of the Deepwater Horizon Disaster
ABRAHAM LUSTGARTEN

W. W. Norton: 2012.
400 pp. \$27.95,
£19.99

Lustgarten reveals holes in BP's. He points to internal and government reports that he says corroborate his claims that BP's cuts led to spills, injuries and deaths. Before Deepwater Horizon, the company twice faced potential debarment from US government contracts, and was convicted of three federal crimes. An analysis by ProPublica, the investigative-journalism newsroom in New York for which Lustgarten reports, revealed that between 1990 and 2009, the US Occupational Safety and Health Administration recorded 518 violations at BP refineries — several times the numbers for comparable companies.

Lustgarten argues that a lax approach to safety permeated BP's actions in the Gulf. The company's spill-preparedness plan seemed to downplay the difficulties of drilling in deep waters, despite fears in the industry that the oil field's unusual geology and extreme pressures and temperatures posed exceptional challenges. In 2000, the

➔ **NATURE.COM**
For Nature's
Deepwater Horizon
special, see:
nature.com/oilspill

US Minerals Management Service (renamed the Bureau of Ocean Energy Management, Regulation and Enforcement (BOEMRE) after the spill), prepared an environmental impact study for Shell, which was also issued as an industry standard. The report warned that a blown-out well in the deep waters of the Gulf might be uncontrollable, and could take months to contain. And BP's senior vice-president for drilling operations in the Gulf resigned from the company months before the spill because he did not believe that BP was committed to safety, according to papers filed in a lawsuit last year.

A BOEMRE report found that Transocean (which owned the drilling rig) and Halliburton (which injected the well's cement casing) were partly culpable for the events that led to the disaster. Ultimately, however, the agency found that BP overrode standard industry practices and made judgement calls that led unintentionally to the blowout.

Lustgarten delves into the Gulf spill itself in only the last two chapters of his book, and handles subsequent events in a brief postscript. That has drawbacks as well as strengths. The environmental costs of the spill and the ongoing scientific inquiries are not explored. The jury is still out on the extent of the environmental damage; for example, it is unclear how marine organisms exposed to oil and dispersants during vulnerable embryonic and larval stages will be affected. Even trace levels of crude oil can alter gene expression in fish, and the widespread, unprecedented use of dispersants below the sea surface may have increased the oil's toxicity to marine organisms. Some scientists say that the full effects will not be understood for years.

Lustgarten also reflects little on the ramifications for public perception of the oil industry. But he does ask why the US government allowed BP to lead the push into deep-water oil exploration. He suggests that the US citizenry's voracious appetite for energy and complacent assumptions about government oversight make it partially culpable.

He ends by noting that little has changed since the disaster. BP's Alaska pipelines are still deteriorating, he suggests. And in the Gulf, drilling resumed in October 2010, after a five-month moratorium on new leases in deep water. US President Barack Obama pledged in his State of the Union address in January to open more than 75% of offshore oil and gas resources to exploration, and environmentalists say that the industry is not making enough progress on safety to merit public trust. It seems that history may well be destined to repeat itself. ■

Amanda Mascarelli is a freelance journalist based in Denver, Colorado, who has written extensively on the oil spill for *Nature*.
e-mail: amanda.mascarelli@gmail.com

Books in brief



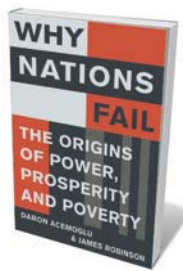
Epigenetics in the Age of Twitter: Pop Culture and Modern Science *Gerald Weissmann* BELLEVUE 300 pp. \$18.95 (2012)

A crackle of erudite energy leaps from this lively commingling of art, culture and science. In 28 essays, biologist Gerald Weissmann explores the complex territory of modern biology and epigenetics in this era of social media. In each, Weissmann finds links between research and elements of history and pop culture, which play off each other to illuminating effect. So US politician Sarah Palin pops up in a discussion of 'Marie Antoinette syndrome', in which hair allegedly whitens overnight; and the 'meltdown' of the mythical Icarus meets the nuclear version at the Fukushima Daiichi power plant in Japan.



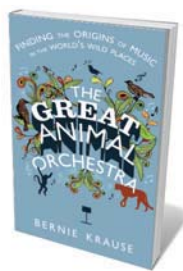
The Idea Factory: Bell Labs and the Great Age of American Innovation *Jon Gertner* PENGUIN 432 pp. \$29.95 (2012)

Lasers, solar cells, prototype mobile phones: from the late 1930s to the mid-1970s, Bell Laboratories — the research arm of US telecommunications giant AT&T — was a powerhouse of innovation. The inventions and ideas emanating from pioneers such as Claude Shannon (information theory) and William Shockley (the transistor) have transformed society. Writer Jon Gertner interviewed employees and researched oral histories to tease out their stories and analyse the organizational ethos that made their achievements possible.



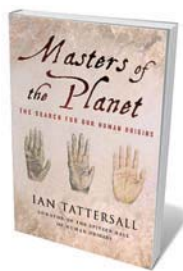
Why Nations Fail: The Origins of Power, Prosperity and Poverty *Daron Acemoglu and James Robinson* CROWN/PROFILE 464 pp. \$30/£25 (2012)

Why is the average US citizen 40 times as prosperous as his or her counterpart in Mali? The cause of such inequity, say economists Daron Acemoglu and James Robinson, is politics. With 15 years of research under their belts, the authors argue that democratized economies and transparent, accountable and responsive governments are the roots of prosperity. Evidence from ancient Rome, the Soviet Union, Europe and the United States makes a compelling case for the power of inclusive institutions to fuel sustainable growth.



The Great Animal Orchestra: Finding the Origins of Music in the World's Wild Places *Bernie Krause* LITTLE, BROWN/PROFILE 288 pp. \$26.99/£12.99 (2012)

Earth, says musician and sound recordist Bernie Krause, pulsates with the clicks, purrs and shrieks of creatures from the yellow-rumped warbler to the snapping shrimp (five times louder than the Grateful Dead, he tells us). Forty years travelling the world to record more than 15,000 species have given Krause a rare insight into the importance of 'biophony': the layered, organized soundscapes of nature. Its disappearance through habitat and species loss is as harmful for human culture and well-being, he says, as it is for ecosystems.



Masters of the Planet: The Search for Our Human Origins *Ian Tattersall* PALGRAVE MACMILLAN 288 pp. £16.99 (2012)

In this succinct and masterful palaeo-chronicle, Ian Tattersall traces how *Homo sapiens* ended up as the world's sole hominin. Tattersall, co-curator of the Spitzer Hall of Human Origins at the American Museum of Natural History in New York, takes us from 6 million years ago in Africa's Rift Valley to the present day. On the way, he brilliantly describes humanity's cousins and rivals, from apes to the other hominins that competed with *H. sapiens* as, tens of thousands of years ago, our ancestors made the cognitive leap to symbolic thought.

US Minerals Management Service (renamed the Bureau of Ocean Energy Management, Regulation and Enforcement (BOEMRE) after the spill), prepared an environmental impact study for Shell, which was also issued as an industry standard. The report warned that a blown-out well in the deep waters of the Gulf might be uncontrollable, and could take months to contain. And BP's senior vice-president for drilling operations in the Gulf resigned from the company months before the spill because he did not believe that BP was committed to safety, according to papers filed in a lawsuit last year.

A BOEMRE report found that Transocean (which owned the drilling rig) and Halliburton (which injected the well's cement casing) were partly culpable for the events that led to the disaster. Ultimately, however, the agency found that BP overrode standard industry practices and made judgement calls that led unintentionally to the blowout.

Lustgarten delves into the Gulf spill itself in only the last two chapters of his book, and handles subsequent events in a brief postscript. That has drawbacks as well as strengths. The environmental costs of the spill and the ongoing scientific inquiries are not explored. The jury is still out on the extent of the environmental damage; for example, it is unclear how marine organisms exposed to oil and dispersants during vulnerable embryonic and larval stages will be affected. Even trace levels of crude oil can alter gene expression in fish, and the widespread, unprecedented use of dispersants below the sea surface may have increased the oil's toxicity to marine organisms. Some scientists say that the full effects will not be understood for years.

Lustgarten also reflects little on the ramifications for public perception of the oil industry. But he does ask why the US government allowed BP to lead the push into deep-water oil exploration. He suggests that the US citizenry's voracious appetite for energy and complacent assumptions about government oversight make it partially culpable.

He ends by noting that little has changed since the disaster. BP's Alaska pipelines are still deteriorating, he suggests. And in the Gulf, drilling resumed in October 2010, after a five-month moratorium on new leases in deep water. US President Barack Obama pledged in his State of the Union address in January to open more than 75% of offshore oil and gas resources to exploration, and environmentalists say that the industry is not making enough progress on safety to merit public trust. It seems that history may well be destined to repeat itself. ■

Amanda Mascarelli is a freelance journalist based in Denver, Colorado, who has written extensively on the oil spill for *Nature*.
e-mail: amanda.mascarelli@gmail.com

Books in brief



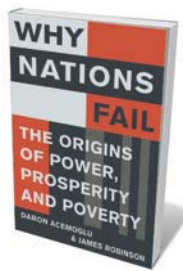
Epigenetics in the Age of Twitter: Pop Culture and Modern Science
Gerald Weissmann BELLEVUE 300 pp. \$18.95 (2012)

A crackle of erudite energy leaps from this lively commingling of art, culture and science. In 28 essays, biologist Gerald Weissmann explores the complex territory of modern biology and epigenetics in this era of social media. In each, Weissmann finds links between research and elements of history and pop culture, which play off each other to illuminating effect. So US politician Sarah Palin pops up in a discussion of 'Marie Antoinette syndrome', in which hair allegedly whitens overnight; and the 'meltdown' of the mythical Icarus meets the nuclear version at the Fukushima Daiichi power plant in Japan.



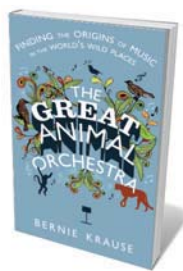
The Idea Factory: Bell Labs and the Great Age of American Innovation
Jon Gertner PENGUIN 432 pp. \$29.95 (2012)

Lasers, solar cells, prototype mobile phones: from the late 1930s to the mid-1970s, Bell Laboratories — the research arm of US telecommunications giant AT&T — was a powerhouse of innovation. The inventions and ideas emanating from pioneers such as Claude Shannon (information theory) and William Shockley (the transistor) have transformed society. Writer Jon Gertner interviewed employees and researched oral histories to tease out their stories and analyse the organizational ethos that made their achievements possible.



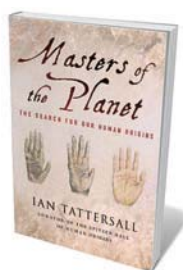
Why Nations Fail: The Origins of Power, Prosperity and Poverty
Daron Acemoglu and James Robinson CROWN/PROFILE 464 pp. \$30/£25 (2012)

Why is the average US citizen 40 times as prosperous as his or her counterpart in Mali? The cause of such inequity, say economists Daron Acemoglu and James Robinson, is politics. With 15 years of research under their belts, the authors argue that democratized economies and transparent, accountable and responsive governments are the roots of prosperity. Evidence from ancient Rome, the Soviet Union, Europe and the United States makes a compelling case for the power of inclusive institutions to fuel sustainable growth.



The Great Animal Orchestra: Finding the Origins of Music in the World's Wild Places
Bernie Krause LITTLE, BROWN/PROFILE 288 pp. \$26.99/£12.99 (2012)

Earth, says musician and sound recordist Bernie Krause, pulsates with the clicks, purrs and shrieks of creatures from the yellow-rumped warbler to the snapping shrimp (five times louder than the Grateful Dead, he tells us). Forty years travelling the world to record more than 15,000 species have given Krause a rare insight into the importance of 'biophony': the layered, organized soundscapes of nature. Its disappearance through habitat and species loss is as harmful for human culture and well-being, he says, as it is for ecosystems.



Masters of the Planet: The Search for Our Human Origins
Ian Tattersall PALGRAVE MACMILLAN 288 pp. £16.99 (2012)

In this succinct and masterful palaeo-chronicle, Ian Tattersall traces how *Homo sapiens* ended up as the world's sole hominin. Tattersall, co-curator of the Spitzer Hall of Human Origins at the American Museum of Natural History in New York, takes us from 6 million years ago in Africa's Rift Valley to the present day. On the way, he brilliantly describes humanity's cousins and rivals, from apes to the other hominins that competed with *H. sapiens* as, tens of thousands of years ago, our ancestors made the cognitive leap to symbolic thought.

PHOTOGRAPHY

Force of nature

Stefan Michalowski and Georgia Smith thrill to artist Berenice Abbott's 'portraits' of physical forces.

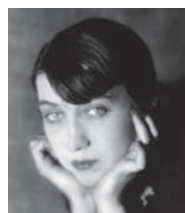
When the Soviet Union launched Sputnik in 1957, the United States was shocked: its rival was taking the lead in science. A group of physicists at the Massachusetts Institute of Technology (MIT) in Cambridge had been pushing for a dramatic upgrade in science education; Sputnik vindicated their efforts. It also prompted a telephone call from a certain Miss Abbott, who said that she had been working on science photography for nearly 20 years. She wanted to add specially created photos to one of the group's projects: a high-school physics textbook encouraging a creative, experiment-based approach in the classroom.

In bringing Berenice Abbott on board, the MIT group unwittingly hired one of the twentieth century's great photographers — who at the time was nearly forgotten, out of fashion and unsure of her next job.

More than 120 of Abbott's photos have gone on display at the Jeu de Paume gallery in Paris, including about two dozen science shots capturing the motion of waves, solid objects and electric currents in striking modernist compositions. (The MIT Museum will

mount its own Abbott exhibition, *Photography and Science: An Essential Unity*, in May.)

Abbott came to photography almost by accident. She had left university for New York in 1918; in 1921, she went to Paris to become a sculptor. Surrealist artist Man Ray hired her as his darkroom assistant and she took to photography instantly, creating iconic portraits of the avant-garde, from James Joyce to Jean Cocteau. She moved back to New York in 1929, eager to document the changing city. In these photographs, her documentary style came into its own. A night shot looking into a forest of brightly lit Manhattan skyscrapers is magnetic; a restaurant window crammed with prices still fascinates. The density of detail — architectural, textural, geometric or social — is striking, and forms a link to the later science photos.



Berenice Abbott (1898–1991),
Photographs
Jeu de Paume, Paris.
Until 29 April 2012.

MAN RAY TRUST/ADAGP

Abbott scoffed at the “arty prettiness” of prominent photographers such as Alfred Stieglitz and Edward Steichen. They rejected her stated ethos of “acute and faithful presentation of what has actually existed in the external world at a particular time and place”.

Next, Abbott seized on science. “There needs to be a friendly interpreter between science and the layman,” she wrote. But funders weren't interested. Steichen's refusal to exhibit her early science photos at New York's Museum of Modern Art killed a potential publishing deal. Abbott studied physics, chemistry and engineering on her own, inventing new techniques and equipment to make her pictures. A stint as photography editor of *Science Illustrated* helped to pay the bills, but Sputnik gave Abbott her big break.

At the Jeu de Paume, one room is filled with her science images. They look abstract, showcasing the geometric lights and darks of high modernism. But when Abbott sets waves in motion in a tray to document interference patterns, or sprinkles iron filings around current-carrying wires, her images reveal the actual forces at play. The finely detailed, composed beauty mesmerises. “We have all seen pretty enough designs abstracted more or less from scientific subjects, but this is mere byplay. This should not be confused with an accurate portrait of scientific phenomena,” Abbott wrote.

Abbott was an inspired choice as a textbook illustrator. One of her best-known works — the strobe-lit arcs of a bouncing golf ball — shows why. An astute student with a ruler could use this elegant image to verify the constant sum of kinetic and potential energies in each arc, and investigating the mysterious disappearance of almost half the kinetic energy of the ball at each bounce would lead to deeper insights. A line drawing wouldn't inspire such investigation, because it could be assumed to be drawn to give the right answer.

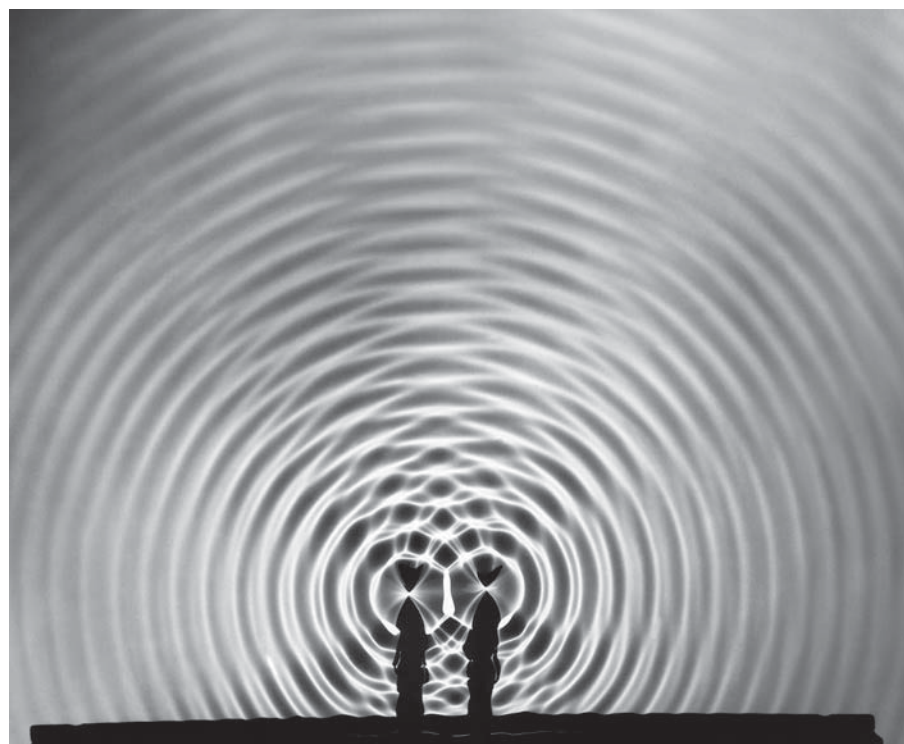
In the textbook, Abbott's images are necessarily small and sometimes murky. This exhibition shows off their beauty, but Abbott as “friendly interpreter” of science might cringe at the absence of explanatory text.

Abbott's three years at MIT were exciting but difficult. She was nearly 60, a plain-spoken artist who lived with another woman, working with world-famous male scientists. She said later that the chauvinistic behaviour of some colleagues had been hard to take. That is easy to believe: as of this writing, the MIT library website attributes the bouncing golf ball photo to one of her male colleagues. ■

Stefan Michalowski is a former particle physicist, and is executive secretary of the OECD Global Science Forum in Paris.

Georgia Smith is a freelance journalist based in Paris.

e-mails: stefanm@noos.fr; georgias@noos.fr



Berenice Abbott's photographs portrayed physical phenomena such as wave interference patterns.

This article does not represent the views of the OECD.

<http://kuai.xunlei.com/d/UQRUEZAIYQHS>



The Olkiluoto 3 nuclear plant in Finland is four years behind schedule and some €2 billion over budget.

The nuclear landscape

The accident at Fukushima has convinced many nations to phase out nuclear power. Economics will be the deciding force, says **Peter Bradford**.

A year after the nuclear meltdowns at the Fukushima Daiichi plant in Japan, tens of thousands of people from the area still cannot return to their homes. Many will not be able to do so during their lifetimes. Only two of Japan's 54 reactors are operating, and there is uncertainty over whether the others will return to service. Once-robust Japanese plans for new nuclear construction have been abandoned, with the government's energy white paper in October 2011 calling for a reduction in reliance on nuclear power.

The disaster has triggered other countries to rethink their nuclear-energy policies. Germany, Belgium and Switzerland have announced plans to close all of their existing reactors (although phase-out plans have been reversed before). They will not build new reactors. Neither will Italy, which closed its reactors after the 1986 Chernobyl accident in what is now Ukraine. In June 2011, some 94% of the Italian electorate voted against any resumption. The Czech Republic has cancelled all but two of its planned plants. Poland has deferred reactor construction for several years. Even France, long a nuclear-power stalwart, may close some reactors early depending on the outcome of elections later this year.

Other nations have made less drastic changes, such as bringing in enhanced safety measures or 'stress tests' for reactors. In countries including the United Kingdom, the United States, China, Russia, India and South Korea, plant closures and cancellations attributed to Fukushima are less likely, but planned expansions will struggle to stay on course.

Before Fukushima, the number of new nuclear plants built in the past 20 years roughly equalled the number retired. The latest re-evaluations suggest at least a short-term downturn for nuclear power's share of global energy production.

Yet the most implacable enemy of nuclear power in the past 30 years has been the risk not to public health, but to investors' wallets. No new nuclear-power project has ever bid successfully in a competitive energy market anywhere in the world.

Fukushima is not the only recent challenge to nuclear power: long-term natural-gas prices and price forecasts have more than

halved in the United States, and globally, electricity demand has dropped on average owing to the economic slowdown. But these events have merely worsened the odds for new nuclear reactors: private capital was no more available before these events than it is now. Accidents at nuclear plants and economic changes have occurred before, within the living memory of global financial communities. The knowledge that any of these events — or new future threats — might happen has long discouraged investors.

Buffeted by economic forces, new nuclear capacity cannot be expected to contribute significantly to global energy supplies or to climate-policy solutions in the decades ahead. Most of the world's reactors are more than 20 years old, so plant completions will be largely offset by retirements. Wise economic and environmental policy will have to allow nuclear energy whatever place it can earn among other energy sources, and not burden it with unnecessary hopes and fears.

THE SECOND COMING

In the United States, where no new reactor has received a construction permit since 1978, a 'second coming' of nuclear construction has been anticipated with as much ballyhoo (and accuracy) as its religious counterpart. In 2009, 31 applications for new reactor projects had been filed or announced. Three years later, only four projects remain on course for construction by 2021. Only on 9 February this year did the US Nuclear Regulatory Commission issue its first licence to build a new reactor in 34 years.

Many people have accepted a version of history in which US nuclear construction wound down because of excessive regulation, environmental alarmism and public fearfulness after the 1979 accident at Three Mile Island in Pennsylvania. For them, the new reactor licence for a site in Georgia proves that US regulators and the public have regained their resolve. But this view rests on a fallacy. Three Mile Island happened at about the same time that competitive market forces were unleashed on the process of power-plant selection. This is the true reason for the decline in reactor construction: some 50 US reactor orders had already been cancelled before the accident.

Chicago-based firm Exelon, the largest US nuclear operator, predicts that electricity from new nuclear plants will cost 12 cents per kilowatt hour (kWh). In comparison, other low-carbon alternatives — combining natural gas with energy efficiency, expansion of existing nuclear plants and wind — will cost 3–11 cents per kWh. Trying to sell new reactors into US power markets that do not value low carbon, and that are predicted to pay no more than 6 cents per kWh for many years,



A YEAR AFTER THE TSUNAMI

For more content, go to:
www.nature.com/japanquake

is a certain route to bankruptcy.

This economic impracticality has been felt globally since the early 1990s. Although the number of nuclear plants in operation jumped from a handful in 1960 to 424 in 1989, they then levelled off, peaking at 444 reactors in 2002 (see 'Not going up'). After Chernobyl, at least a dozen plants of the same design were taken out of action. Older, smaller plants were decommissioned. But the determining factor was the drop in construction in the United States and western Europe — a result of the cost of building new plants and the refusal of investors to bear the risks of cancellations, cost overruns and the emergence of cheaper alternatives.

At the time of the Fukushima disaster, only four countries (China, Russia, India and South Korea) were building more than two reactors. In these four nations, citizens pay for the new reactors the government chooses to build through direct subsidies or energy price hikes. In a few cases, such as the Olkiluoto 3 reactor under construction in Finland, the company providing the reactor promises a fixed price or looks to its own government for finance. Olkiluoto is being built by Areva, an entity largely owned by the French government, which was gambling that the project would jump-start demand for its newest reactor design. As Olkiluoto is four years behind schedule and more than €2 billion (US\$2.7 billion) over budget, that gamble has fallen flat.

The only nations for which building high-cost nuclear plants has any prospect of becoming competitive are those that combine very limited access to natural gas with high growth (Japan once among them). Even then, there is no reason to pay over the odds for nuclear. Instead, all greenhouse-gas mitigating measures and alternative-energy sources should have to compete with each other in the marketplace.

"It would be ideal if Fukushima could steer us away from prophecies and towards a sensible assessment."

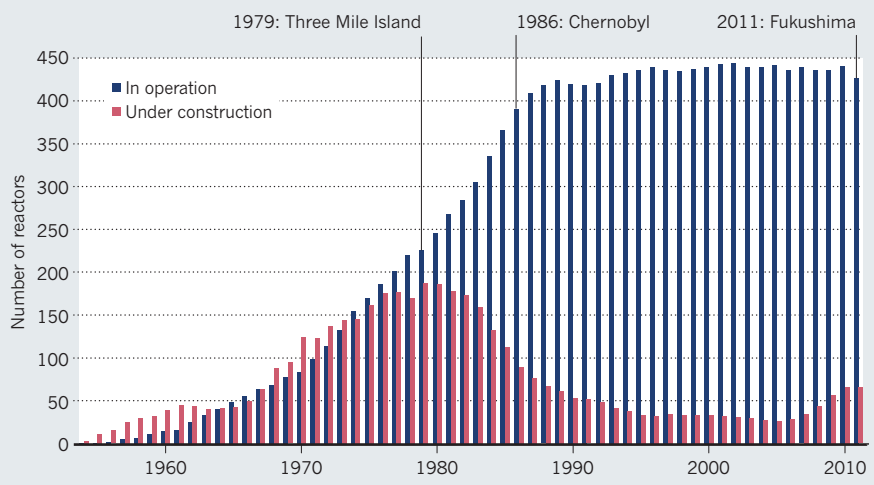
POLITICAL BALANCE

On top of the ongoing dire economic situation, Fukushima has shifted the political equation. The Chernobyl disaster was widely seen as a product of the lax Soviet safety culture, in which an unsafe design was combined with reckless operation. But Japan was considered to be in the technological and regulatory front rank. Champions of nuclear power in the US Congress (and even President Barack Obama on one occasion) had held up Japan as a nuclear success story that US regulators and energy policy-makers should strive to emulate.

After Chernobyl, Italy and Germany

NOT GOING UP

Building of new nuclear reactors declined after the 1979 Three Mile Island accident, as economic realities hit home. The number in operation has levelled out since the late 1980s.



decided to phase out nuclear energy, as Sweden had done after Three Mile Island. All three later reconsidered as concerns over climate change and years without a conspicuous reactor accident shifted the political balance back towards nuclear energy. The unfavourable economics did not change, but the willingness of governments to override economics with subsidies and mandated purchases did.

In the United States, this change has produced a paradox. The right-wing legislators who swept into state and federal office in 2010 have conflicting tendencies towards nuclear energy: they are largely pro-nuclear, pro-market, climate-change sceptics, yet they disparage governmental backing of particular technologies with taxpayers' money. Before Fukushima, the net effect seemed to be towards state and federal measures that were favourable for new nuclear capacity, such as increases in taxpayer-backed loan guarantees or laws that made energy customers responsible for runaway costs or cancelled plants. After Fukushima, those measures quietly died.

In the United Kingdom, too, a conservative government pledged to produce new nuclear power without government subsidies. It is now having to torture the language of new policies to subsidize new reactors without this being recognized as such.

It is impossible to say what will happen next for nuclear energy. The indications from emerging carbon markets are that they will not produce a high enough price to support new nuclear. These markets are not providing enough emissions avoidance either, so the price for carbon could go higher. If Exelon's predictions are correct, and if gas prices remain low, carbon prices would have to more than triple to make new nuclear look economical. At that level,

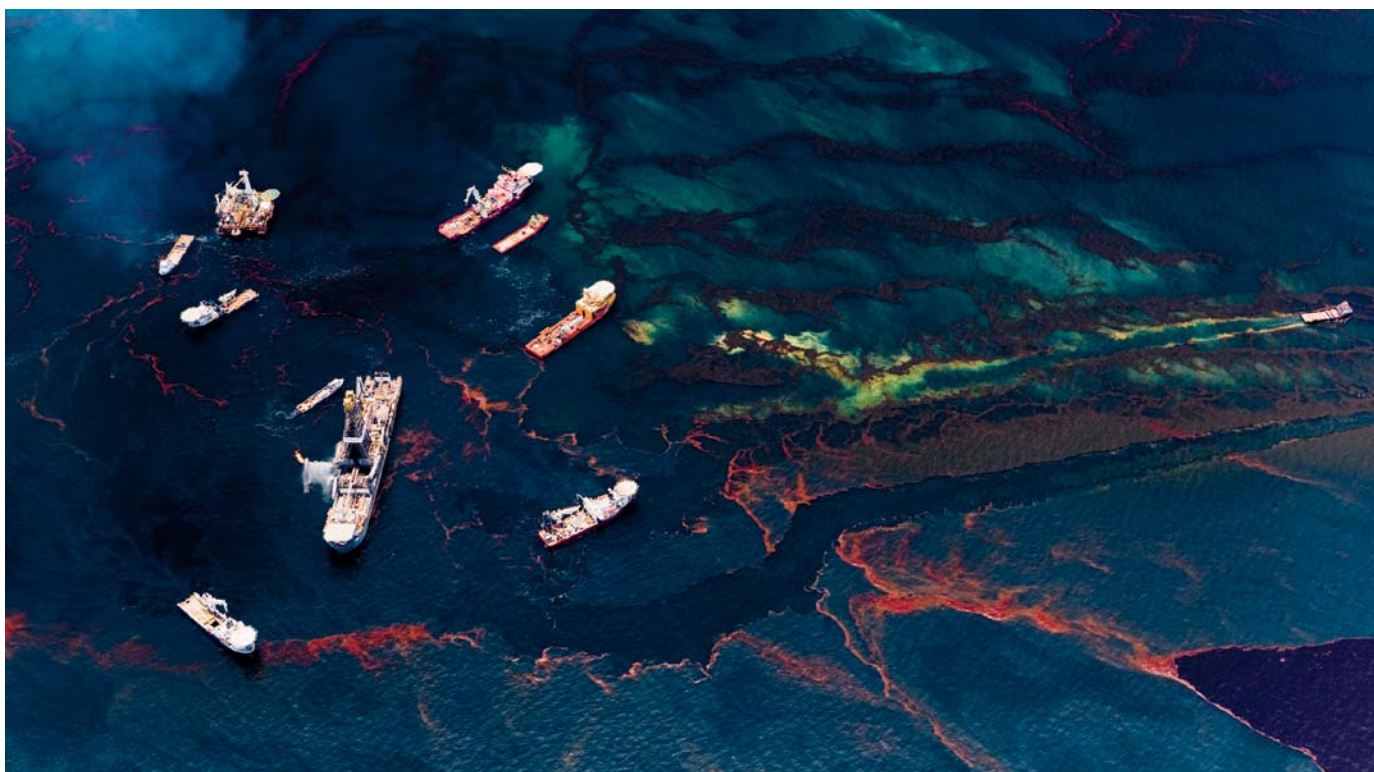
many other alternatives could be built on a large scale more quickly.

The hopes of the US nuclear industry now rest on the new reactor in Georgia — one of the few states with a law assuring that energy customers will pay all reactor costs regardless of price. Energy secretary Steven Chu, visiting the site last month, intoned: "If this project goes forward and is built on-budget, on-time and on-schedule, that would be a very good thing. A lot of other companies will say, 'OK. We now know we can do this and it would be a good investment.'" But companies will say no such thing unless they can also produce energy at a competitive price. Instead, they will join Exelon in saying 'we can do this and go broke'.

Many countries state the unknowable or the implausible with great certainty. Germany has declared that existing nuclear capacity is too dangerous and can be phased out at little cost. The Middle East and the Czech Republic assert that new nuclear is a vital low-carbon energy source. Others, including the United Kingdom, have said that there will soon be new nuclear without subsidies.

It would be ideal if Fukushima could steer us away from these prophecies and towards a sensible assessment of market economics, climate science and nuclear risks. Then nuclear power would serve the public, not the other way around. I don't know how many reactors we would get, but we would get the number that we need. ■

Peter Bradford is an adjunct professor teaching nuclear power and public policy at Vermont Law School, South Royalton, Vermont 05068, USA. He served on the US Nuclear Regulatory Commission in 1977–82. e-mail: perubrad@aol.com



D. BELTRA/GREENPEACE/NATUREPL.COM

Oil on the waters of the Gulf of Mexico could alter gene expression in fish and other marine organisms for years to come.

ENERGY

Plumbing the depths

A chronicle of events preceding the Deepwater Horizon oil spill in the Gulf of Mexico has a thriller-like edge, finds **Amanda Mascarelli**.

Many people have wondered how the Deepwater Horizon oil spill of 2010 happened, given the technological know-how and industry back-up systems in place to prevent such a catastrophe. In *Run To Failure*, journalist Abraham Lustgarten explains how, in his view, the blowout was not only predictable, but inevitable. The book is one of the few that focus on the roots of the disaster — and is refreshingly different from the many looking mainly at the immediate tragedy and its aftermath.

Run to Failure reads like a thriller, complete with whistle-blowers and double agents. On the basis of meticulous investigative reporting during and after the spill, analysis of BP documents and e-mails, and interviews with current and former industry insiders, Lustgarten paints a picture of neglect, hollow proclamations about safety and environmental stewardship, and draconian cost-trimming going back two decades. That history, he believes, helped to pave the way to the disaster.

The Gulf of Mexico has long been seen

as the frontier of oil exploration. In the mid-1990s, most of the oil extracted in the region came from shallow waters. But deep-water wells are much more bountiful. To push into the depths, BP invested tens of billions of dollars in the Gulf, says Lustgarten, and sent in its most technologically advanced drilling rigs and resources. However, to subsidize its investment, Lustgarten alleges, BP made cuts elsewhere, including at its ageing US refineries and on the Alaska North Slope — which had been the source of much of BP's oil production since the 1970s, but where reserves were dwindling.

The oil industry as a whole has a patchy record on safety and the environment, and



Run To Failure: BP and the Making of the Deepwater Horizon Disaster
ABRAHAM LUSTGARTEN
W. W. Norton: 2012.
400 pp. \$27.95,
£19.99

Lustgarten reveals holes in BP's. He points to internal and government reports that he says corroborate his claims that BP's cuts led to spills, injuries and deaths. Before Deepwater Horizon, the company twice faced potential debarment from US government contracts, and was convicted of three federal crimes. An analysis by ProPublica, the investigative-journalism newsroom in New York for which Lustgarten reports, revealed that between 1990 and 2009, the US Occupational Safety and Health Administration recorded 518 violations at BP refineries — several times the numbers for comparable companies.

Lustgarten argues that a lax approach to safety permeated BP's actions in the Gulf. The company's spill-preparedness plan seemed to downplay the difficulties of drilling in deep waters, despite fears in the industry that the oil field's unusual geology and extreme pressures and temperatures posed exceptional challenges. In 2000, the

➔ **NATURE.COM**
For Nature's
Deepwater Horizon
special, see:
nature.com/oilspill

US Minerals Management Service (renamed the Bureau of Ocean Energy Management, Regulation and Enforcement (BOEMRE) after the spill), prepared an environmental impact study for Shell, which was also issued as an industry standard. The report warned that a blown-out well in the deep waters of the Gulf might be uncontrollable, and could take months to contain. And BP's senior vice-president for drilling operations in the Gulf resigned from the company months before the spill because he did not believe that BP was committed to safety, according to papers filed in a lawsuit last year.

A BOEMRE report found that Transocean (which owned the drilling rig) and Halliburton (which injected the well's cement casing) were partly culpable for the events that led to the disaster. Ultimately, however, the agency found that BP overrode standard industry practices and made judgement calls that led unintentionally to the blowout.

Lustgarten delves into the Gulf spill itself in only the last two chapters of his book, and handles subsequent events in a brief postscript. That has drawbacks as well as strengths. The environmental costs of the spill and the ongoing scientific inquiries are not explored. The jury is still out on the extent of the environmental damage; for example, it is unclear how marine organisms exposed to oil and dispersants during vulnerable embryonic and larval stages will be affected. Even trace levels of crude oil can alter gene expression in fish, and the widespread, unprecedented use of dispersants below the sea surface may have increased the oil's toxicity to marine organisms. Some scientists say that the full effects will not be understood for years.

Lustgarten also reflects little on the ramifications for public perception of the oil industry. But he does ask why the US government allowed BP to lead the push into deep-water oil exploration. He suggests that the US citizenry's voracious appetite for energy and complacent assumptions about government oversight make it partially culpable.

He ends by noting that little has changed since the disaster. BP's Alaska pipelines are still deteriorating, he suggests. And in the Gulf, drilling resumed in October 2010, after a five-month moratorium on new leases in deep water. US President Barack Obama pledged in his State of the Union address in January to open more than 75% of offshore oil and gas resources to exploration, and environmentalists say that the industry is not making enough progress on safety to merit public trust. It seems that history may well be destined to repeat itself. ■

Amanda Mascarelli is a freelance journalist based in Denver, Colorado, who has written extensively on the oil spill for *Nature*.
e-mail: amanda.mascarelli@gmail.com

Books in brief



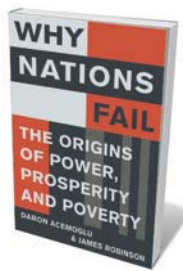
Epigenetics in the Age of Twitter: Pop Culture and Modern Science
Gerald Weissmann BELLEVUE 300 pp. \$18.95 (2012)

A crackle of erudite energy leaps from this lively commingling of art, culture and science. In 28 essays, biologist Gerald Weissmann explores the complex territory of modern biology and epigenetics in this era of social media. In each, Weissmann finds links between research and elements of history and pop culture, which play off each other to illuminating effect. So US politician Sarah Palin pops up in a discussion of 'Marie Antoinette syndrome', in which hair allegedly whitens overnight; and the 'meltdown' of the mythical Icarus meets the nuclear version at the Fukushima Daiichi power plant in Japan.



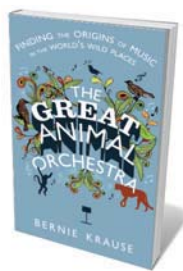
The Idea Factory: Bell Labs and the Great Age of American Innovation
Jon Gertner PENGUIN 432 pp. \$29.95 (2012)

Lasers, solar cells, prototype mobile phones: from the late 1930s to the mid-1970s, Bell Laboratories — the research arm of US telecommunications giant AT&T — was a powerhouse of innovation. The inventions and ideas emanating from pioneers such as Claude Shannon (information theory) and William Shockley (the transistor) have transformed society. Writer Jon Gertner interviewed employees and researched oral histories to tease out their stories and analyse the organizational ethos that made their achievements possible.



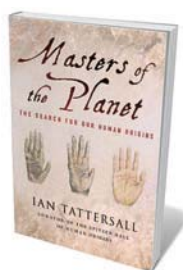
Why Nations Fail: The Origins of Power, Prosperity and Poverty
Daron Acemoglu and James Robinson CROWN/PROFILE 464 pp. \$30/£25 (2012)

Why is the average US citizen 40 times as prosperous as his or her counterpart in Mali? The cause of such inequity, say economists Daron Acemoglu and James Robinson, is politics. With 15 years of research under their belts, the authors argue that democratized economies and transparent, accountable and responsive governments are the roots of prosperity. Evidence from ancient Rome, the Soviet Union, Europe and the United States makes a compelling case for the power of inclusive institutions to fuel sustainable growth.



The Great Animal Orchestra: Finding the Origins of Music in the World's Wild Places
Bernie Krause LITTLE, BROWN/PROFILE 288 pp. \$26.99/£12.99 (2012)

Earth, says musician and sound recordist Bernie Krause, pulsates with the clicks, purrs and shrieks of creatures from the yellow-rumped warbler to the snapping shrimp (five times louder than the Grateful Dead, he tells us). Forty years travelling the world to record more than 15,000 species have given Krause a rare insight into the importance of 'biophony': the layered, organized soundscapes of nature. Its disappearance through habitat and species loss is as harmful for human culture and well-being, he says, as it is for ecosystems.



Masters of the Planet: The Search for Our Human Origins
Ian Tattersall PALGRAVE MACMILLAN 288 pp. £16.99 (2012)

In this succinct and masterful palaeo-chronicle, Ian Tattersall traces how *Homo sapiens* ended up as the world's sole hominin. Tattersall, co-curator of the Spitzer Hall of Human Origins at the American Museum of Natural History in New York, takes us from 6 million years ago in Africa's Rift Valley to the present day. On the way, he brilliantly describes humanity's cousins and rivals, from apes to the other hominins that competed with *H. sapiens* as, tens of thousands of years ago, our ancestors made the cognitive leap to symbolic thought.



Q&A Vijay Iyer

Jazz experimentalist

*Vijay Iyer is a New York jazz pianist who has academic roots in physics and music cognition. As he releases *Accelerando* — a follow-up to his 2009 world number one jazz album *Historicity* — he talks about the bodily origins of rhythm, the science of improvisation and the social function of music.*

How did you start out?

My father, a pharmaceutical chemist, was in the first wave of Indian immigrants to the United States. I was raised to value science, and I had a few piano lessons as a teenager. At Yale University in New Haven, Connecticut, where I studied physics and mathematics, I played music the whole time, but never considered it as a career. Then, as soon as I arrived at the University of California, Berkeley, to do a PhD in physics, I won a jazz-piano competition. I was pursuing research in solid-state physics, modelling defects in magnetic materials, but music was turning into a part-time job. Finally it took over.

How did you begin to study music perception?

At 24, I hit a wall with research and started taking courses outside physics. One was with David Wessel at Berkeley, who did important early research on timbre perception and computer music. I was ready to become an artist, but he helped me to create my own PhD programme in music perception and

Accelerando

VIJAY IYER
ACT Music + Vision:
2012. \$14.90

subjects in the psychology literature are wealthy undergraduates at liberal-arts institutions. It is very hard to tease out the cognitive universals of music from a sample of white, suburban teenagers listening to Mozart.

What was your dissertation about?

I described cognition as intimately tied to the senses, situated and embodied in the body's many timescales. The breath has a flexible period of several seconds that corresponds to the lengths of typical musical phrases. The heartbeat and footsteps occur once or twice a second, close to the natural pulse of most music, and speech rhythms correspond pretty well to finger motions and to rapid rhythmic detail.

Why focus on rhythm?

At the primal level, music is rhythm first, the sound of bodies in synchronous action.

cognition. I began to see that the field can be dangerously monocultural. Most test sub-

That is why there is a pulse in music. Rhythm perception is an imagined movement in the motor centres of the brain. Our skill for coordinating our actions is the real foundation of music, and possibly of civilization.

What kind of music research would you like to see?

I would like to see researchers collaborating with musicians, which doesn't happen much. Charles Limb, a surgeon at Johns Hopkins University in Baltimore, Maryland, does brain scans of people improvising on a piano. He has also taken scans of rappers freestyling. His findings — that improvisers turn down the part of the brain that governs inhibitions and turn up the part that governs creativity — are specific and plausible.

How do your research and music feed one another?

For a while, I tried to keep up with research conferences. But making music is its own kind of research. The musicians that I like are experimenting constantly, testing ideas with an audience night after night, and trying to push the boundaries of what they know.

How do you use scientific ideas in your music?

Some composers might write a string quartet 'about' string theory, but that is just inspiration, it is not really discovery. I'm more of an experimentalist. There is an auditory illusion of a constantly ascending pitch, known as Shepard tones: the musical equivalent of M. C. Escher's infinite staircase. As the pitch goes up, the distribution of harmonics shifts down, and your ear can't find the place where it doubles back on itself. I used this illusion in a string quartet by asking the players to perform a synchronized glissando in parallel octaves and imposing a bell curve on their amplitudes. It worked. After that, I asked, can we do this with tempo? At the end of the title track on *Historicity*, there is a rhythm that constantly decelerates. On *Accelerando*, there is a piece giving the illusion of constant acceleration, of a tempo that flexes.

What is the future of music?

People walk around with headphones on, thinking of music as a solitary, personalized pursuit. But it has connected us by synchronizing our actions throughout human history. Because we are so engrossed in the technical aspects, it is easy for scientists, and even for musicians, to forget that the effects of music are primarily emotional. That is why people keep it in their lives. ■

INTERVIEW BY JASCHA HOFFMAN

Correspondence

Step up funding for flu prevention

Experts on pandemic influenza need to convince governments and the public that, although the 2009–10 'swine flu' pandemic proved to be relatively mild, the threat posed by future flu pandemics is severe and warrants more investment in public education and research (*Nature* **482**, 131; 2012).

Without effective education, there is a danger that the low mortality in the 2009–10 pandemic will exacerbate public scepticism over the lethality of flu. This could undermine compliance with basic preventative measures, such as conscientious personal hygiene and social isolation.

Research funding for pandemic flu prevention and treatment must be stepped up. The US government spent US\$26.2 billion on HIV/AIDS activities in 2010, a year in which 1.8 million people died from AIDS worldwide. Tuberculosis killed a comparable number in 2009 and received \$224 million for research from the US National Institutes of Health alone in 2010. A pandemic strain of flu could, in theory, cause up to 60 million deaths worldwide in just two years. The \$627 million that the US government spent on flu research in 2008–10 (see go.nature.com/c2wwke) would then seem decidedly insufficient.

Simon N. Williams

Northwestern University, Chicago, Illinois, USA.
simonwilliams@northwestern.edu

Keep jellyfish numbers in check

It may be unclear whether jellyfish numbers are rising globally (*Nature* **482**, 20–21; 2012), but this should not distract us from taking urgent action to control populations in those degraded ecosystems where particular species have undeniably increased.

You mention the “paradigm” that jellyfish are increasing globally, but this consensus view is not held by the scientific community because jellyfish time series are scarce. Even with new database syntheses of scientific data, supplemented by anecdotal information from fishermen, news reports, historical research cruises and modelled reconstructions of historic ecosystems, we are unlikely to reach a consensus in the near future.

There is compelling evidence that some jellyfish species pose a risk in particular marine systems, so we believe that precautionary action should be taken now. Efforts could focus on increasing surveillance (currently scanty for jellyfish) and minimizing habitat eutrophication, overfishing and species translocation, all thought to cause jellyfish outbreaks.

Anthony J. Richardson CSIRO Marine and Atmospheric Research, Dutton Park, Queensland; and University of Queensland, St Lucia, Australia.
anthony.richardson@csiro.au

Daniel Pauly University of British Columbia, Vancouver, Canada.

Mark J. Gibbons University of the Western Cape, Bellville, South Africa.

Clinical standards not practical in the lab

Gholson Lyon calls for more rigorous standards in genetics research so that results can be disclosed to subjects and their families as valid clinical diagnostic information (*Nature* **482**, 300–301; 2012). This well-intentioned proposal is impractical because it would add huge costs for only occasional benefit.

Lyon is essentially recommending that the standards of Good Clinical Practice (see go.nature.com/1pado7) should apply to laboratory research as well. The cost and complexity of carrying out all research to

comply with these regulations would be crippling. Compliance would demand validated environments for sample-tracking (Lyon's focus) and for all analytical work. In whole-genome sequencing, validation for error-free determination of those 3 billion or so data points in the human genome would be almost impossible.

A more pragmatic alternative would be to incorporate specific provisions in the informed consent for people wanting to access their data. This would carry the disclaimer that results are not formally validated for clinical use, which would require independent replication to meet the standards of Good Clinical Practice. This re-testing could be made available at no cost to the patient should the clinical need arise.

The cost of responding to such an occasional requirement would be insignificant in comparison with processing every subject's specimens under regulatory-compliant conditions.

Klaus Lindpaintner SDIX, Newark, Delaware, USA; and Bio-Banks Consulting, Basel, Switzerland.

klindpaintner@sdix.com

Sugar: an excess of anything can harm

As director-general of the World Sugar Research Organisation, I wish to point out some shortcomings in the latest discussion of sugar's impact on health (*Nature* **482**, 27–29; 2012).

Robert Lustig and colleagues incorrectly say that sugar consumption has tripled worldwide since the 1960s. The global population has more than doubled in that time, so the increase in sugar supply per head is more like 60%. In fact, the United States, Canada and the United Kingdom all show only marginal changes over the past few decades in average sugar consumption as

a proportion of food-energy intake.

The authors argue that sugar can kill because of its supposed influence on metabolic syndrome (itself a controversial concept), indirectly implicating a WHO Technical Report that draws no such conclusion. There is little consistent effect on the symptoms of this syndrome in people who eat up to three times more sugar than the average Western intake (A. S. Truswell *Am. J. Clin. Nutr.* **59**, 710S–718S; 1994). Neither have any deaths been attributed to dietary sugars in an exhaustive analysis of US mortality figures (G. Danaei *et al. PLoS Medicine* **6**, e100058; 2009).

The Food and Agriculture Organization of the United Nations, the US Food and Nutrition Board, and the European Food Standards Authority have all considered the issues now revisited by Lustig *et al.* and find no reliable evidence that typical sugar consumption contributes to any disease apart from dental caries. Without evidence that reducing sugar consumption would improve public health, Lustig and colleagues' policy proposals are irrelevant.

Scientific controversies should be settled by consideration of all the available evidence, not of a seemingly biased selection. Overconsumption of anything is harmful, including of water and air.

Richard C. Cottrell World Sugar Research Organisation, London, UK. rcottrell@wsro.org
Competing financial interests declared (see go.nature.com/ek29gr).

CONTRIBUTIONS

Correspondence may be sent to correspondence@nature.com after consulting the guidelines at <http://go.nature.com/cmchno>. Alternatively, readers may comment online: www.nature.com/nature.

Human reproductive assistance

What is the biological explanation for menopause, and for female survival beyond it? A study suggests that competition for help in ancestral societies may have been key to the evolution of this unusual human trait.

KIM HILL & A. MAGDALENA HURTADO

Humans stand out spectacularly from other species. We have language and cumulative culture, and we cooperate in large groups of unrelated individuals. We also have a distinctive life history, characterized by long juvenile dependence and menopause — the cessation of female reproductive function that occurs only about two-thirds of the way through the protracted adult lifespan. Recent research has examined how ancestral social structure and interaction networks may help to explain our cooperative tendencies^{1,2}. According to a study by Mace and Alvergne³ published in *Proceedings of the Royal Society*, these social tendencies may also have contributed to the evolution of menopause, by channelling competition for assistance among women.

The saying goes that it takes a village to raise a child (Fig. 1). And, indeed, our ancestors evolved as cooperative breeders^{4,5}, dependent on helpers for successful reproduction. In these circumstances, adults must weigh up the relative costs and benefits of themselves reproducing or acting as helpers. Kin helpers must decide which relative to help in order to most efficiently promote the spread of their shared genes, and non-kin helpers must consider how help might influence opportunities for mating or other benefits. Thus, the combination of kin and non-kin available to provide assistance influences the reproductive patterns in a community.

In most vertebrate cooperative breeders, such as meerkats or magpies, younger females forgo reproduction to help older females reproduce. In humans, however, the pattern is reversed. To assess whether competition for help among people may explain this profound difference, Mace and Alvergne studied historical demographic data (obtained between 1949 and 1975) from farming villages in Gambia. In these villages, extended families resided in a single compound, with males remaining in their birth compound and adult females moving to others to marry, a situation termed virilocality. It is reasonable to hypothesize that menopause evolved within a virilocal social system, because African apes also show a virilocal pattern of female dispersal and so,



Figure 1 | Competition breeds success. Mace and Alvergne's study³ of demographic data from women (such as those shown here) in some Gambian villages shows that competition for assistance from kin and non-kin is a key contributor to female reproductive success, and may help to explain the evolution of menopause.

presumably, did our earliest human ancestors⁶. Even if this were not the case, we know that menopause has been maintained for thousands of years in a world of predominantly virilocal human societies, such that virilocality and menopause must be compatible under the pressure of natural selection.

There is an asymmetry of relatedness in virilocal societies, in which the older women of a compound are genetically related to the offspring of their son's wives, but younger immigrant women do not share genes with their mother-in-law's children. If women compete for the assistance that they can give each other, the asymmetry of relatedness to each other's offspring might explain why older women are most willing to forgo reproduction to help⁷. However, Mace and Alvergne's data analysis³ contradicts this hypothesis. They show that it is in fact the reproductive rate of older women that improves as the number of competing women in a compound increases, whereas

reproduction in younger women declines.

A simple explanation for this trend could be that older women, whom Mace and Alvergne show are more closely related to other adults in the compound, have a greater influence on the distribution of goods and services. Yet when the authors analysed the data further to take into account mother–daughter pairs, they found the reverse trend, with an older woman's reproductive output decreasing when her daughter begins to reproduce. This effect occurs even if the daughter no longer lives in the same compound. Although the authors acknowledge that daughters often do not disperse far (staying within the same village, for example), such that dispersed individuals may still help each other, the complexity of findings returns the focus of menopause research to the previously popular idea of mother–daughter helping.

Under the 'grandmother hypothesis', menopause is theorized to have evolved mainly so that mothers would aid their own daughters' reproduction. This hypothesis stems from the idea that the greater food-provisioning capacity of older women in ancestral societies might have had sufficient impact on their daughters' fertility and grandchildren's survival to favour the evolution of reproductive cessation in order to make women available to help⁸. But although the helping effects of grandmothers are often statistically significant, studies show that they are probably not great enough to outweigh the loss of genetic contribution that could otherwise be attained through direct reproduction⁹. Furthermore, the grandmother hypothesis does not explain how menopause has been maintained in virilocal populations. Mace and Alvergne's finding that the effects of competition on reproductive success in overlapping generations extend to women who do not reside in the same household helps to resolve this potential inconsistency.

However, to understand how menopause is maintained in societies in which daughters disperse far away, other social interactions must be considered. Women in ancestral societies often lived with kin of both sexes¹, such that reproductive-age females competed for subsidies offered by pair-bonded mates, kin of various ages and males hoping to entice future mating. Mothers and daughters were probably

C. THEGE/ALAMY

not each other's main food provisioners⁴, so an analysis of female reproductive success must also take into account help from others, especially provisioning males.

How do outside helpers choose where to invest their help? We suggest that the answer may come, perhaps surprisingly, from social insects. Some species of insects that live in colonies can detect an individual's relative fertility from the smell of pheromones and other chemical secretions¹⁰. They then adjust their behaviour, such that low-fertility females are not allowed to breed¹¹ and low-fertility queens are replaced¹² by more-fertile breeders. One species even shows the typical human pattern — all young colony females are reproductively active and then progress to a post-reproductive worker stage as they age¹³.

Perhaps similar responses evolved in ancestral human societies, and ageing females with detectably decreasing fertility were less favoured by helpers and potential investing mates in the social group. Withdrawal of food provisioning might then have amplified the age-specific decline in fertility from slight to extreme — a possible explanation for the precipitous decline of a woman's fertility with age, which occurs long before her children reproduce, and to an extent that is far more pronounced than the impact of daughters' reproduction, as measured by Mace and Alvergne. Under this scenario, evolutionary pressure would have favoured older females helping descendants, rather than trying to reproduce without help, thereby promoting the evolution and maintenance of menopause.

Although it is not yet definitively clear why menopause evolved, Mace and Alvergne's findings show that competition for help probably had a key role — another example of how complexity in ancestral human social structure was involved in the evolution of critical human traits. It is exciting to ponder on which other intriguing, and exceptional, aspects of human biology may be more completely understood by considering the social elements of our history. ■

Kim Hill is in the School of Human Evolution and Social Change, and the Institute of Human Origins, and **A. Magdalena Hurtado** is in the School of Human Evolution and Social Change, Arizona State University, Tempe, Arizona 85287-2402, USA.
e-mails: kim.hill@asu.edu; amhurtad@asu.edu

- Hill, K. *et al.* *Science* **331**, 1286–1288 (2011).
- Apicella, C. L., Marlowe, F. W., Fowler, J. H. & Christakis, N. A. *Nature* **481**, 497–501 (2012).
- Mace, R. & Alvergne, A. *Proc. R. Soc. B* <http://dx.doi.org/10.1098/rspb.2011.2424> (2012).
- Hill, K. & Hurtado, A. M. *Proc. R. Soc. B* **276**, 3863–3870 (2009).
- Hrdy, S. B. *Mother and Others: The Evolutionary Origins of Mutual Understanding* (Harvard Univ. Press, 2009).
- Chapais, B. *Primeval Kinship: How Pair-Bonding Gave Birth to Human Society* (Harvard Univ. Press, 2008).

- Johnstone, R. A. & Kant, M. A. *Proc. R. Soc. B* **277**, 3765–3771 (2010).
- Hawkes, K., O'Connell, J. F., Blurton Jones, N. G., Alvarez, H. & Charnov, E. L. *Proc. Natl Acad. Sci. USA* **95**, 1336–1339 (1998).
- Hill, K. & Hurtado, A. M. *Ache Life History: The Ecology and Demography of a Foraging People* (Aldine de Gruyter, 1996).

- Liebig, J., Peters, C., Oldham, N., Marcstädter, C. & Hölldobler, B. *Proc. Natl Acad. Sci. USA* **97**, 4124–4131 (2000).
- Cuvillier-Hot, V., Lenoir, A. & Peeters, C. *Behav. Ecol.* **15**, 970–975 (2004).
- Seeley, T. D. *Honeybee Ecology* (Princeton Univ. Press, 1985).
- Tsuji, K. *Anim. Behav.* **39**, 843–849 (1990).

PHYSICAL CHEMISTRY

Single molecules filmed dancing on a table top

By ripping an electron away from a molecule and then slamming it back again, the motion of nuclei in a molecule has been tracked with extremely high temporal and spatial resolution. SEE LETTER P.194

MISHA Y. IVANOV

In 1909, the physicists Hans Geiger and Ernest Marsden shot α -particles (helium nuclei) at a thin gold foil¹. Following the suggestion of Ernest Rutherford, they were looking for (and found) a substantial number of particles that were deflected by the foil at angles larger than 90° — rather than passing straight through it, as ideas about the structure of atoms predicted. According to Rutherford², the discovery “was quite the most incredible event that has ever happened to me in my life. It was almost as incredible as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you.” Rutherford concluded³ from these results that atoms have very compact nuclei, where all of their positive charge is concentrated, surrounded by much more loosely spaced electrons. The α -particles rebounded because they collided head-on with fully charged gold nuclei.

Ever since then, the scattering of charged particles has been used to study the ångström-scale structure of matter, effectively providing a range of ‘cameras’ for taking pictures of the microcosm. Reporting on page 194 of this issue, Blaga *et al.*⁴ describe an unusual variant of these cameras that has allowed them to study the motion of nuclei in single molecules of oxygen and nitrogen at sub-ångström spatial resolution, and with a temporal resolution of just a few femtoseconds (1 femtosecond is 10⁻¹⁵ s).

For molecules to be seen at their ångström-scale dimensions, they must be illuminated with a beam of some sort that has an ångström-scale wavelength. Pulses of X-rays are an option, but they must be very short and exceptionally bright to take snapshots of quickly moving nuclei in single molecules. Free-electron lasers can produce such X-ray pulses, but they require massive facilities that cost billions of dollars to build and maintain.

Short electron pulses are a much cheaper alternative to X-rays, and can be made using table-top-sized apparatus. Electrons are, of course, particles, but quantum mechanics tells us that they also have wave-like properties. When accelerated to about 3% of the speed of light, their wavelength becomes short enough to image objects at the ångström scale. What's more, because electrons are charged and have low mass, accelerating them with electric fields is quite straightforward. Electrons also scatter well, much better than X-rays, and thus provide bright signals for imaging. Extremely short pulses of electrons can therefore take stroboscopic ‘snapshots’ of moving nuclei in molecules, much like the frames of a movie. Indeed, this is the basis of the techniques of ultrafast electron diffraction⁵ and ultrafast electron microscopy⁶.

The charged nature of electrons, however, also presents a problem: repelled by their like charges, electrons do not want to stay together in groups. Ultrashort electron pulses containing many electrons therefore tend to become much longer by the time they reach their target during imaging experiments, reducing the temporal resolution that can be achieved. With very few electrons per pulse (ideally, just one), the pulses can stay short — but then the strength of the imaging signal plummets, making the observation of individual molecules exceptionally difficult. This is why ultrafast electron diffraction and ultrafast electron microscopy have failed to take freeze-frame pictures of individual molecules at a temporal resolution of a few femtoseconds. Blaga *et al.*⁴ have now accomplished this feat, remarkably using just one electron per pulse.

The authors' ultrafast electron camera is based on an unusual approach that was originally proposed about ten years ago^{7–9}: the electron that is used to illuminate a molecule is pulled out from the molecule itself by a strong

not each other's main food provisioners⁴, so an analysis of female reproductive success must also take into account help from others, especially provisioning males.

How do outside helpers choose where to invest their help? We suggest that the answer may come, perhaps surprisingly, from social insects. Some species of insects that live in colonies can detect an individual's relative fertility from the smell of pheromones and other chemical secretions¹⁰. They then adjust their behaviour, such that low-fertility females are not allowed to breed¹¹ and low-fertility queens are replaced¹² by more-fertile breeders. One species even shows the typical human pattern — all young colony females are reproductively active and then progress to a post-reproductive worker stage as they age¹³.

Perhaps similar responses evolved in ancestral human societies, and ageing females with detectably decreasing fertility were less favoured by helpers and potential investing mates in the social group. Withdrawal of food provisioning might then have amplified the age-specific decline in fertility from slight to extreme — a possible explanation for the precipitous decline of a woman's fertility with age, which occurs long before her children reproduce, and to an extent that is far more pronounced than the impact of daughters' reproduction, as measured by Mace and Alvergne. Under this scenario, evolutionary pressure would have favoured older females helping descendants, rather than trying to reproduce without help, thereby promoting the evolution and maintenance of menopause.

Although it is not yet definitively clear why menopause evolved, Mace and Alvergne's findings show that competition for help probably had a key role — another example of how complexity in ancestral human social structure was involved in the evolution of critical human traits. It is exciting to ponder on which other intriguing, and exceptional, aspects of human biology may be more completely understood by considering the social elements of our history. ■

Kim Hill is in the School of Human Evolution and Social Change, and the Institute of Human Origins, and **A. Magdalena Hurtado** is in the School of Human Evolution and Social Change, Arizona State University, Tempe, Arizona 85287-2402, USA.
e-mails: kim.hill@asu.edu; amhurtad@asu.edu

- Hill, K. *et al.* *Science* **331**, 1286–1288 (2011).
- Apicella, C. L., Marlowe, F. W., Fowler, J. H. & Christakis, N. A. *Nature* **481**, 497–501 (2012).
- Mace, R. & Alvergne, A. *Proc. R. Soc. B* <http://dx.doi.org/10.1098/rspb.2011.2424> (2012).
- Hill, K. & Hurtado, A. M. *Proc. R. Soc. B* **276**, 3863–3870 (2009).
- Hrdy, S. B. *Mother and Others: The Evolutionary Origins of Mutual Understanding* (Harvard Univ. Press, 2009).
- Chapais, B. *Primeval Kinship: How Pair-Bonding Gave Birth to Human Society* (Harvard Univ. Press, 2008).

- Johnstone, R. A. & Kant, M. A. *Proc. R. Soc. B* **277**, 3765–3771 (2010).
- Hawkes, K., O'Connell, J. F., Blurton Jones, N. G., Alvarez, H. & Charnov, E. L. *Proc. Natl Acad. Sci. USA* **95**, 1336–1339 (1998).
- Hill, K. & Hurtado, A. M. *Ache Life History: The Ecology and Demography of a Foraging People* (Aldine de Gruyter, 1996).

- Liebig, J., Peters, C., Oldham, N., Marcstädter, C. & Hölldobler, B. *Proc. Natl Acad. Sci. USA* **97**, 4124–4131 (2000).
- Cuvillier-Hot, V., Lenoir, A. & Peeters, C. *Behav. Ecol.* **15**, 970–975 (2004).
- Seeley, T. D. *Honeybee Ecology* (Princeton Univ. Press, 1985).
- Tsuji, K. *Anim. Behav.* **39**, 843–849 (1990).

PHYSICAL CHEMISTRY

Single molecules filmed dancing on a table top

By ripping an electron away from a molecule and then slamming it back again, the motion of nuclei in a molecule has been tracked with extremely high temporal and spatial resolution. SEE LETTER P.194

MISHA Y. IVANOV

In 1909, the physicists Hans Geiger and Ernest Marsden shot α -particles (helium nuclei) at a thin gold foil¹. Following the suggestion of Ernest Rutherford, they were looking for (and found) a substantial number of particles that were deflected by the foil at angles larger than 90° — rather than passing straight through it, as ideas about the structure of atoms predicted. According to Rutherford², the discovery “was quite the most incredible event that has ever happened to me in my life. It was almost as incredible as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you.” Rutherford concluded³ from these results that atoms have very compact nuclei, where all of their positive charge is concentrated, surrounded by much more loosely spaced electrons. The α -particles rebounded because they collided head-on with fully charged gold nuclei.

Ever since then, the scattering of charged particles has been used to study the ångström-scale structure of matter, effectively providing a range of ‘cameras’ for taking pictures of the microcosm. Reporting on page 194 of this issue, Blaga *et al.*⁴ describe an unusual variant of these cameras that has allowed them to study the motion of nuclei in single molecules of oxygen and nitrogen at sub-ångström spatial resolution, and with a temporal resolution of just a few femtoseconds (1 femtosecond is 10⁻¹⁵ s).

For molecules to be seen at their ångström-scale dimensions, they must be illuminated with a beam of some sort that has an ångström-scale wavelength. Pulses of X-rays are an option, but they must be very short and exceptionally bright to take snapshots of quickly moving nuclei in single molecules. Free-electron lasers can produce such X-ray pulses, but they require massive facilities that cost billions of dollars to build and maintain.

Short electron pulses are a much cheaper alternative to X-rays, and can be made using table-top-sized apparatus. Electrons are, of course, particles, but quantum mechanics tells us that they also have wave-like properties. When accelerated to about 3% of the speed of light, their wavelength becomes short enough to image objects at the ångström scale. What's more, because electrons are charged and have low mass, accelerating them with electric fields is quite straightforward. Electrons also scatter well, much better than X-rays, and thus provide bright signals for imaging. Extremely short pulses of electrons can therefore take stroboscopic ‘snapshots’ of moving nuclei in molecules, much like the frames of a movie. Indeed, this is the basis of the techniques of ultrafast electron diffraction⁵ and ultrafast electron microscopy⁶.

The charged nature of electrons, however, also presents a problem: repelled by their like charges, electrons do not want to stay together in groups. Ultrashort electron pulses containing many electrons therefore tend to become much longer by the time they reach their target during imaging experiments, reducing the temporal resolution that can be achieved. With very few electrons per pulse (ideally, just one), the pulses can stay short — but then the strength of the imaging signal plummets, making the observation of individual molecules exceptionally difficult. This is why ultrafast electron diffraction and ultrafast electron microscopy have failed to take freeze-frame pictures of individual molecules at a temporal resolution of a few femtoseconds. Blaga *et al.*⁴ have now accomplished this feat, remarkably using just one electron per pulse.

The authors' ultrafast electron camera is based on an unusual approach that was originally proposed about ten years ago^{7–9}: the electron that is used to illuminate a molecule is pulled out from the molecule itself by a strong



50 Years Ago

From the fermentation broth of a strain of *Fusidium* a hitherto unrecorded antibiotic, for which the name 'fusidic acid' is proposed, has been isolated ... The activity was determined by the agar cup-plate method using *Staphylococcus aureus* as test organism ...

From the concentrated aqueous solution obtained, fusidic acid was precipitated as a crystalline benzene solvate on acidification in the presence of benzene ... Fusidic acid is non-toxic. The subcutaneous and oral LD_{50} in mice were found to be 1.2 gm. and 1.5 gm. per kgm. body-weight, respectively ... Daily oral administration of fusidic acid to rats in doses of 0.4 gm. per kgm. body-weight over a period of 6 months was well-tolerated. Post-mortem examination revealed no pathological changes.

From *Nature* 10 March 1962

100 Years Ago

Mr. G. R. M. Temple sends us from York a copy of a photograph ... which illustrates very clearly the result of the expansion of water by freezing during the recent frost. The bottle was filled with clean water and tightly corked; when the water had frozen a stem of ice about $4\frac{1}{2}$ in. in length was found to be projecting from the bottle ... This stem represented, of course, the increase of volume undergone by the water in passing from the liquid to the solid state. The bottle must have been cracked while solidification was going on, otherwise the water would have escaped.

From *Nature* 7 March 1912

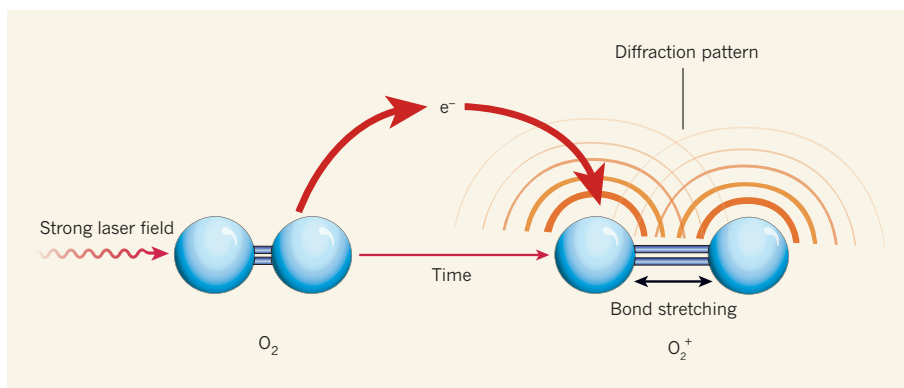


Figure 1 | Molecular movies. Blaga *et al.*⁴ have used laser-induced electron diffraction to 'film' single molecules in motion. In the authors' technique, a molecule such as oxygen is irradiated with a strong infrared laser field, pulling away an electron (e^-) and causing the bond between the atoms in the molecule to stretch. The electron is accelerated in the laser field, then directed back towards the resulting ion. Collision of the electron with nuclei in the ion causes the electron to rebound and to generate a diffraction pattern — an electron spectrum. This spectrum contains information about the positions of the nuclei at the moment of collision. By performing a series of experiments in which the time delay between the electron's departure from, and collision with, the molecule is varied, a series of snapshots (spectra) of the molecular motion can be taken, much like the frames of a movie.

infrared laser field (Fig. 1). Thanks to the high intensity of the laser field (its field strength reaches several volts per ångström), a distance of less than 100 Å is sufficient to accelerate the electrons to several per cent of light speed. And because of the oscillating nature of the field, the accelerated electron can then be turned around and slammed back into the parent molecule.

The removal of an electron from a molecule sets the molecule's nuclei in motion. Subsequent re-collision of the electron with the molecule takes a snapshot of what has happened since the electron was taken away. This snapshot takes the form of an electron spectrum — the patterns of minimum and maximum electron intensity that are generated by the diffraction of electron waves by nuclei, not unlike the patterns of ripples formed when water waves are disrupted by rocks.

Even with relatively few molecules in the focus of the laser beam, there is a good chance that the removed electron will hit the target, because it never leaves the vicinity of the parent molecule. This ensures a strong imaging signal. Furthermore, everything happens in a time roughly equal to the period of the laser field's oscillation cycle (a few femtoseconds for the radiation used in Blaga and colleagues' experiment), ensuring high temporal resolution.

To make any movie, snapshots must be taken at different time delays after the motion has begun. In Blaga and colleagues' case, the delay is that between the time the electron is pulled from a molecule and the time it returns. This delay is determined by the length of the laser cycle, and so the authors simply repeated their experiment at different wavelengths to obtain a series of frames at different times. Just like Rutherford before them, they observed electrons scattering at large angles as they hit nuclei almost head-on. The scattering angles

and the diffraction patterns created by scattered electrons sensitively depend on the locations of the nuclei — even if the positions of the nuclei shifted by only a fraction of an ångström, the difference was clearly visible in the experiment.

So far, so good. But Blaga *et al.* had to overcome many challenges to reconstruct molecular motion from the measured electron signals. The main difficulty arose from the presence of the strong laser field — ironically, the same field that is integral to the inner workings of their camera. In their experiments, information about molecular structure is encoded in the observed electron spectra. But because each electron is accelerated by the laser field after scattering, the patterns become distorted. Extracting images from the electron spectra therefore requires an accurate theoretical analysis⁷.

The physics of laser-driven electron-molecule re-collision is rich and complex. The next step will be to perform a detailed analysis of all of the similarly complex information encoded in Blaga and colleagues' electron spectra, to deal with various side effects introduced by the strong laser field, and to extend the scope of the technique to larger molecules. This will involve a lot of work, but obtaining the ability to film the motion of electrons, holes (quasiparticles generated by the absence of electrons) and nuclei in isolated molecules, with sub-femtosecond, sub-ångström resolution, is worth the effort. No doubt Blaga *et al.* would agree with Lenin¹⁰ that "for us, cinematography is the most important form of art". ■

Misha Y. Ivanov is in the Department of Physics, Imperial College London, London SW7 2AZ, UK, and at the Max Born Institute for Nonlinear Optics and the Department of Physics, Humboldt University, Berlin, Germany.
e-mail: m.ivanov@imperial.ac.uk

1. Geiger, H. & Marsden, E. *Proc. R. Soc. A* **82**, 495–500 (1909).
2. Cassidy, D., Holton, G. & Rutherford, J. *Understanding Physics* 632 (Birkhäuser, 2002).
3. Rutherford, E. *Phil. Mag. Ser. 6* **21**, 669–688 (1911).
4. Blaga, C. I. *et al. Nature* **483**, 194–197 (2012).
5. Scialini, G. & Miller, R. J. D. *Rep. Prog. Phys.* **74**, 096101 (2011).
6. Zewail, A. H. *Annu. Rev. Phys. Chem.* **57**, 65–103 (2006).
7. Lein, M., Marangos, J. P. & Knight, P. L. *Phys. Rev. A* **66**, 051404 (2002).
8. Spanner, M., Smirnova, O., Corkum, P. B. & Ivanov, M. Y. *J. Phys. B* **37**, L243–L250 (2004).
9. Yurchenko, S. N., Patchkovskii, S., Litvinyuk, I. V., Corkum, P. B. & Yudin, G. L. *Phys. Rev. Lett.* **93**, 223003 (2004).
10. Lenin, V. I. *Complete Works* 5th edn, Vol. 44, 579 (Political Literature Publishing House, 1970).

CELL BIOLOGY

The sensation of stretch

Piezo proteins have been shown to form large ion channels that serve a sensory function in fruitflies. The findings help to explain how Piezos convert mechanical force into biological signals. SEE ARTICLE P.176 & LETTER P.209

PHILIP A. GOTTLIEB & FREDERICK SACHS

An organism's ability to react to mechanical stimuli is crucial for its survival. One set of biological tools for sensing mechanical stress is the ion channels that open in response to tension in cell membranes, often called mechanosensitive channels (MSCs). Much progress has been made by studying MSCs in different systems¹, but one of the biggest breakthroughs came in 2010, when Ardem Patapoutian's group² identified a cation-selective MSC that responds directly to mechanical forces in the membrane of certain mouse cells. They found that two similar proteins — Piezo1 and Piezo2 — each can form MSCs in different cell types.

In two papers^{3,4} published in this issue, Patapoutian's group presents another milestone in our understanding of mechanical transduction. The team reports that Piezo proteins form channels composed of four large identical subunits, and that the expression of these channels is directly related to nociceptive responses — neural processes associated with potentially harmful stimuli — in the larvae of the fruitfly *Drosophila melanogaster*. This is the first time that the detailed biophysical properties of a cation-selective MSC have been correlated with changes in behaviour.

In mice, Piezo1 contains about 2,500 amino acids and is arranged into more than 30 transmembrane domains — making it structurally different from other known ion channels. Patapoutian and colleagues previously reported² that the expression of Piezo1-encoding genes in various mechanically insensitive cells made those cells sensitive to mechanical stimuli. Furthermore, the conductance and inactivation properties of Piezo1 are similar to those of the first MSCs to be identified⁵, which were found in non-sensory cells. Piezo1 is also the first MSC from a eukaryote (organisms such as plants and animals) that is known⁶ to be inhibited by the peptide GsMTx4, a compound widely used as a channel blocker

in the study of MSCs. Patapoutian's group now asks whether Piezo1 is itself an ion channel, or whether it modifies the activity of another channel (or another protein).

In the first of the two papers (page 176), Coste *et al.*³ convincingly argue that Piezo1 proteins assemble to form a tetramer, on the basis of results from two complementary methods. In the first approach, the authors attached a green fluorescent protein to Piezo1. They then used light to extinguish the fluorescence of the resulting construct, and observed the loss of fluorescence using single-molecule imaging techniques. The fluorescence diminished in four quantized steps, suggesting that Piezo1 assembles as a tetramer.

Coste and colleagues' second approach was to chemically crosslink the subunits of Piezo1. When the authors subjected the crosslinked sample to electrophoresis, they observed discrete bands on the electrophoretic sizing gel that could be explained by the formation of a tetramer. The team also used mass spectroscopy to show that no other proteins are associated with Piezo1, which suggests that Piezo1 does not exert its effects by modifying the activity of another protein. Whether the tetramer is indeed the functionally active channel formed by Piezo1 remains to be seen.

The authors found³ that the tetrameric complex has a molecular mass of about 1.2 million daltons, that it has 120–160 transmembrane segments and that the monomer is different from those of other known channels. The large size of the tetramer is not obviously advantageous for mechanical activation, because structural changes associated with the activation of other MSCs are known to be small⁷ (about an ångström). Furthermore, bacterial MSCs are highly sensitive to membrane tension⁸, despite being considerably smaller than Piezo channels. The unusual architecture of the Piezo1 complex therefore indicates that we have more to learn about this protein family.

Coste *et al.* went on to demonstrate that purified Piezo1 could be reconstituted in

planar lipid bilayers and liposomes (artificial vesicles made from lipid bilayers), and that these reconstituted proteins had conductance properties characteristic of a cation-selective ion channel. This means that auxiliary force-coupling structures, such as the cytoskeleton of cells, are not required to activate Piezo1 in membranes — although the authors' experiments did not prove that the reconstituted proteins were mechanosensitive.

In the second paper (page 209), Kim *et al.*⁴ focus on a Piezo protein, DmPiezo, in *D. melanogaster*. Like Piezo1 and Piezo2 in mice, the authors found that DmPiezo responds to mechanical stimuli when expressed in human cells. When the researchers knocked out the *Dmpiezo* gene from *Drosophila* larvae, the larvae's behavioural response to noxious mechanical stimuli was reduced compared with that of wild-type larvae, although their responses to other mechanical stimuli, such as gentle touch, were unaffected.

Similarly, by specifically depleting the levels of DmPiezo in the sensory neurons used for nociception in larvae, Kim *et al.* diminished the animals' response to noxious mechanical stimuli. This effect could be reversed by reintroducing DmPiezo into the larvae. However, knocking out *Dmpiezo* did not completely abolish the nociceptive response, suggesting the presence of parallel signalling pathways for mechanosensitivity in the larvae. When the authors knocked out both *Dmpiezo* and *pickpocket* (a gene that encodes another type of ion channel), they observed complete loss of nociception.

Mouse and *Drosophila* Piezo proteins share some characteristics — they exhibit similar mechanical sensitivity and time-dependent inactivation, for example. But there are also differences: mouse Piezo1 has a higher ion conductance than DmPiezo, and is more sensitive to ruthenium red, a compound used to block the pores of the transient receptor potential (TRP) family of ion channels. The reactivity of ruthenium red with Piezo1 is a reminder that the compound cannot be used solely as a TRP channel inhibitor.

The study by Kim *et al.*⁴ suggests that Piezo proteins are a new family of eukaryotic mechanosensitive channels. Perhaps the most pleasing aspect of their work, however, is the demonstration of a relationship between mechanical transduction and sensory processing: if force is applied to a cell containing DmPiezo, an influx of positive ions through the channel makes the cell interior more positive. The resulting change in potential across the membrane signals to the animal that a noxious stimulus is present. What could be simpler? ■

Philip A. Gottlieb and Frederick Sachs are at the Center for Single Molecule Biophysics, Department of Physiology and Biophysics, School of Medicine and Biomedical Sciences,



50 Years Ago

From the fermentation broth of a strain of *Fusidium* a hitherto unrecorded antibiotic, for which the name 'fusidic acid' is proposed, has been isolated ... The activity was determined by the agar cup-plate method using *Staphylococcus aureus* as test organism ...

From the concentrated aqueous solution obtained, fusidic acid was precipitated as a crystalline benzene solvate on acidification in the presence of benzene ... Fusidic acid is non-toxic. The subcutaneous and oral LD_{50} in mice were found to be 1.2 gm. and 1.5 gm. per kgm. body-weight, respectively ... Daily oral administration of fusidic acid to rats in doses of 0.4 gm. per kgm. body-weight over a period of 6 months was well-tolerated. Post-mortem examination revealed no pathological changes.

From *Nature* 10 March 1962

100 Years Ago

Mr. G. R. M. Temple sends us from York a copy of a photograph ... which illustrates very clearly the result of the expansion of water by freezing during the recent frost. The bottle was filled with clean water and tightly corked; when the water had frozen a stem of ice about $4\frac{1}{2}$ in. in length was found to be projecting from the bottle ... This stem represented, of course, the increase of volume undergone by the water in passing from the liquid to the solid state. The bottle must have been cracked while solidification was going on, otherwise the water would have escaped.

From *Nature* 7 March 1912

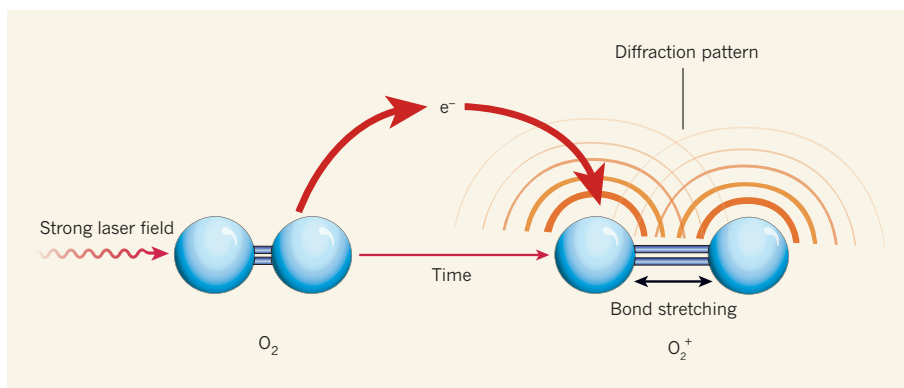


Figure 1 | Molecular movies. Blaga *et al.*⁴ have used laser-induced electron diffraction to 'film' single molecules in motion. In the authors' technique, a molecule such as oxygen is irradiated with a strong infrared laser field, pulling away an electron (e^-) and causing the bond between the atoms in the molecule to stretch. The electron is accelerated in the laser field, then directed back towards the resulting ion. Collision of the electron with nuclei in the ion causes the electron to rebound and to generate a diffraction pattern — an electron spectrum. This spectrum contains information about the positions of the nuclei at the moment of collision. By performing a series of experiments in which the time delay between the electron's departure from, and collision with, the molecule is varied, a series of snapshots (spectra) of the molecular motion can be taken, much like the frames of a movie.

infrared laser field (Fig. 1). Thanks to the high intensity of the laser field (its field strength reaches several volts per ångström), a distance of less than 100 Å is sufficient to accelerate the electrons to several per cent of light speed. And because of the oscillating nature of the field, the accelerated electron can then be turned around and slammed back into the parent molecule.

The removal of an electron from a molecule sets the molecule's nuclei in motion. Subsequent re-collision of the electron with the molecule takes a snapshot of what has happened since the electron was taken away. This snapshot takes the form of an electron spectrum — the patterns of minimum and maximum electron intensity that are generated by the diffraction of electron waves by nuclei, not unlike the patterns of ripples formed when water waves are disrupted by rocks.

Even with relatively few molecules in the focus of the laser beam, there is a good chance that the removed electron will hit the target, because it never leaves the vicinity of the parent molecule. This ensures a strong imaging signal. Furthermore, everything happens in a time roughly equal to the period of the laser field's oscillation cycle (a few femtoseconds for the radiation used in Blaga and colleagues' experiment), ensuring high temporal resolution.

To make any movie, snapshots must be taken at different time delays after the motion has begun. In Blaga and colleagues' case, the delay is that between the time the electron is pulled from a molecule and the time it returns. This delay is determined by the length of the laser cycle, and so the authors simply repeated their experiment at different wavelengths to obtain a series of frames at different times. Just like Rutherford before them, they observed electrons scattering at large angles as they hit nuclei almost head-on. The scattering angles

and the diffraction patterns created by scattered electrons sensitively depend on the locations of the nuclei — even if the positions of the nuclei shifted by only a fraction of an ångström, the difference was clearly visible in the experiment.

So far, so good. But Blaga *et al.* had to overcome many challenges to reconstruct molecular motion from the measured electron signals. The main difficulty arose from the presence of the strong laser field — ironically, the same field that is integral to the inner workings of their camera. In their experiments, information about molecular structure is encoded in the observed electron spectra. But because each electron is accelerated by the laser field after scattering, the patterns become distorted. Extracting images from the electron spectra therefore requires an accurate theoretical analysis⁷.

The physics of laser-driven electron-molecule re-collision is rich and complex. The next step will be to perform a detailed analysis of all of the similarly complex information encoded in Blaga and colleagues' electron spectra, to deal with various side effects introduced by the strong laser field, and to extend the scope of the technique to larger molecules. This will involve a lot of work, but obtaining the ability to film the motion of electrons, holes (quasiparticles generated by the absence of electrons) and nuclei in isolated molecules, with sub-femtosecond, sub-ångström resolution, is worth the effort. No doubt Blaga *et al.* would agree with Lenin¹⁰ that "for us, cinematography is the most important form of art". ■

Misha Y. Ivanov is in the Department of Physics, Imperial College London, London SW7 2AZ, UK, and at the Max Born Institute for Nonlinear Optics and the Department of Physics, Humboldt University, Berlin, Germany.
e-mail: m.ivanov@imperial.ac.uk

1. Geiger, H. & Marsden, E. *Proc. R. Soc. A* **82**, 495–500 (1909).
2. Cassidy, D., Holton, G. & Rutherford, J. *Understanding Physics* 632 (Birkhäuser, 2002).
3. Rutherford, E. *Phil. Mag. Ser. 6* **21**, 669–688 (1911).
4. Blaga, C. I. *et al. Nature* **483**, 194–197 (2012).
5. Scialini, G. & Miller, R. J. D. *Rep. Prog. Phys.* **74**, 096101 (2011).
6. Zewail, A. H. *Annu. Rev. Phys. Chem.* **57**, 65–103 (2006).
7. Lein, M., Marangos, J. P. & Knight, P. L. *Phys. Rev. A* **66**, 051404 (2002).
8. Spanner, M., Smirnova, O., Corkum, P. B. & Ivanov, M. Y. *J. Phys. B* **37**, L243–L250 (2004).
9. Yurchenko, S. N., Patchkovskii, S., Litvinyuk, I. V., Corkum, P. B. & Yudin, G. L. *Phys. Rev. Lett.* **93**, 223003 (2004).
10. Lenin, V. I. *Complete Works* 5th edn, Vol. 44, 579 (Political Literature Publishing House, 1970).

CELL BIOLOGY

The sensation of stretch

Piezo proteins have been shown to form large ion channels that serve a sensory function in fruitflies. The findings help to explain how Piezos convert mechanical force into biological signals. [SEE ARTICLE P.176](#) & [LETTER P.209](#)

PHILIP A. GOTTLIEB & FREDERICK SACHS

An organism's ability to react to mechanical stimuli is crucial for its survival. One set of biological tools for sensing mechanical stress is the ion channels that open in response to tension in cell membranes, often called mechanosensitive channels (MSCs). Much progress has been made by studying MSCs in different systems¹, but one of the biggest breakthroughs came in 2010, when Ardem Patapoutian's group² identified a cation-selective MSC that responds directly to mechanical forces in the membrane of certain mouse cells. They found that two similar proteins — Piezo1 and Piezo2 — each can form MSCs in different cell types.

In two papers^{3,4} published in this issue, Patapoutian's group presents another milestone in our understanding of mechanical transduction. The team reports that Piezo proteins form channels composed of four large identical subunits, and that the expression of these channels is directly related to nociceptive responses — neural processes associated with potentially harmful stimuli — in the larvae of the fruitfly *Drosophila melanogaster*. This is the first time that the detailed biophysical properties of a cation-selective MSC have been correlated with changes in behaviour.

In mice, Piezo1 contains about 2,500 amino acids and is arranged into more than 30 transmembrane domains — making it structurally different from other known ion channels. Patapoutian and colleagues previously reported² that the expression of Piezo1-encoding genes in various mechanically insensitive cells made those cells sensitive to mechanical stimuli. Furthermore, the conductance and inactivation properties of Piezo1 are similar to those of the first MSCs to be identified⁵, which were found in non-sensory cells. Piezo1 is also the first MSC from a eukaryote (organisms such as plants and animals) that is known⁶ to be inhibited by the peptide GsMTx4, a compound widely used as a channel blocker

in the study of MSCs. Patapoutian's group now asks whether Piezo1 is itself an ion channel, or whether it modifies the activity of another channel (or another protein).

In the first of the two papers (page 176), Coste *et al.*³ convincingly argue that Piezo1 proteins assemble to form a tetramer, on the basis of results from two complementary methods. In the first approach, the authors attached a green fluorescent protein to Piezo1. They then used light to extinguish the fluorescence of the resulting construct, and observed the loss of fluorescence using single-molecule imaging techniques. The fluorescence diminished in four quantized steps, suggesting that Piezo1 assembles as a tetramer.

Coste and colleagues' second approach was to chemically crosslink the subunits of Piezo1. When the authors subjected the crosslinked sample to electrophoresis, they observed discrete bands on the electrophoretic sizing gel that could be explained by the formation of a tetramer. The team also used mass spectroscopy to show that no other proteins are associated with Piezo1, which suggests that Piezo1 does not exert its effects by modifying the activity of another protein. Whether the tetramer is indeed the functionally active channel formed by Piezo1 remains to be seen.

The authors found³ that the tetrameric complex has a molecular mass of about 1.2 million daltons, that it has 120–160 transmembrane segments and that the monomer is different from those of other known channels. The large size of the tetramer is not obviously advantageous for mechanical activation, because structural changes associated with the activation of other MSCs are known to be small⁷ (about an ångström). Furthermore, bacterial MSCs are highly sensitive to membrane tension⁸, despite being considerably smaller than Piezo channels. The unusual architecture of the Piezo1 complex therefore indicates that we have more to learn about this protein family.

Coste *et al.* went on to demonstrate that purified Piezo1 could be reconstituted in

planar lipid bilayers and liposomes (artificial vesicles made from lipid bilayers), and that these reconstituted proteins had conductance properties characteristic of a cation-selective ion channel. This means that auxiliary force-coupling structures, such as the cytoskeleton of cells, are not required to activate Piezo1 in membranes — although the authors' experiments did not prove that the reconstituted proteins were mechanosensitive.

In the second paper (page 209), Kim *et al.*⁴ focus on a Piezo protein, DmPiezo, in *D. melanogaster*. Like Piezo1 and Piezo2 in mice, the authors found that DmPiezo responds to mechanical stimuli when expressed in human cells. When the researchers knocked out the *Dmpiezo* gene from *Drosophila* larvae, the larvae's behavioural response to noxious mechanical stimuli was reduced compared with that of wild-type larvae, although their responses to other mechanical stimuli, such as gentle touch, were unaffected.

Similarly, by specifically depleting the levels of DmPiezo in the sensory neurons used for nociception in larvae, Kim *et al.* diminished the animals' response to noxious mechanical stimuli. This effect could be reversed by reintroducing DmPiezo into the larvae. However, knocking out *Dmpiezo* did not completely abolish the nociceptive response, suggesting the presence of parallel signalling pathways for mechanosensitivity in the larvae. When the authors knocked out both *Dmpiezo* and *pickpocket* (a gene that encodes another type of ion channel), they observed complete loss of nociception.

Mouse and *Drosophila* Piezo proteins share some characteristics — they exhibit similar mechanical sensitivity and time-dependent inactivation, for example. But there are also differences: mouse Piezo1 has a higher ion conductance than DmPiezo, and is more sensitive to ruthenium red, a compound used to block the pores of the transient receptor potential (TRP) family of ion channels. The reactivity of ruthenium red with Piezo1 is a reminder that the compound cannot be used solely as a TRP channel inhibitor.

The study by Kim *et al.*⁴ suggests that Piezo proteins are a new family of eukaryotic mechanosensitive channels. Perhaps the most pleasing aspect of their work, however, is the demonstration of a relationship between mechanical transduction and sensory processing: if force is applied to a cell containing DmPiezo, an influx of positive ions through the channel makes the cell interior more positive. The resulting change in potential across the membrane signals to the animal that a noxious stimulus is present. What could be simpler? ■

Philip A. Gottlieb and Frederick Sachs are at the Center for Single Molecule Biophysics, Department of Physiology and Biophysics, School of Medicine and Biomedical Sciences,

State University of New York at Buffalo,
Buffalo, New York 14214, USA.
e-mails: philgott@buffalo.edu; sachs@buffalo.edu

1. Chalfie, M. *Nature Rev. Mol. Cell. Biol.* **10**, 44–52 (2009).
2. Coste, B. *et al. Science* **330**, 55–60 (2010).
3. Coste, B. *et al. Nature* **483**, 176–181 (2012).
4. Kim, S. E., Coste, B., Chadha, A., Cook, B. & Patapoutian, A. *Nature* **483**, 209–212 (2012).
5. Guharay, F. & Sachs, F. *J. Physiol. (Lond.)* **352**, 685–701 (1984).
6. Bae, C., Sachs, F. & Gottlieb, P. A. *Biochemistry* **50**, 6295–6300 (2011).
7. Honoré, E., Patel, A. J., Chemin, J., Suchyna, T. & Sachs, F. *Proc. Natl Acad. Sci. USA* **103**, 6859–6864 (2006).
8. Kung, C., Martinac, B. & Sukharev, S. *Annu. Rev. Microbiol.* **64**, 313–329 (2010).

GENOMICS

Gorilla gorilla gorilla

The gorilla genome reveals that genetic similarities among humans and the apes are more complex than expected, and allows a fresh assessment of the evolutionary mechanisms that led to the primate species seen today. [SEE ARTICLE P.169](#)

RICHARD A. GIBBS & JEFFREY ROGERS

Humans and the apes are the living representatives of the superfamily Hominoidea, which also contains many extinct species. Deciphering the evolutionary relationships between these species is an essential step in our understanding of the biological richness of the planet, and of our own evolutionary history. The draft assembly of the whole-genome sequence of a female western lowland gorilla (*Gorilla gorilla gorilla*) named Kamilah (Fig. 1), presented by Scally *et al.*¹ on page 169 of this issue, provides insight into how a single hominid lineage separated into the extant human, chimpanzee and gorilla branches.

The authors' genome assembly, made possible by the advent of cheaper next-generation sequencing methods, is a much-anticipated

addition to more than five-year-old Sanger sequence data on the gorilla genome. The latest assembly, like other contemporary mammalian genome data sets (except for human and mouse), has some gaps and shortcomings. Yet it is revelatory. The standard view of the primate evolutionary tree is that chimpanzees and humans share a more recent common ancestor with each other than either shares with gorillas. Accordingly, the most closely related sequence for any human gene should be found in the chimpanzee. However, Scally and colleagues' demonstrate that, although this is true for most genes, large fractions of the ape genomes contradict this simple pattern.

Molecular phylogenetics uses comparative analyses of DNA sequences to determine relatedness among species. Usually, a single individual of each species is sequenced, so genetic diversity within a species is overlooked.

When the time between successive evolutionary branch points, or speciation events, is relatively long, the between-species genetic differences that accumulate after a single lineage divides into two descendent branches will stand out against the background of within-species variation. But if two or more evolutionary divergence events occur close together in time, the genetic variation present in the last common ancestor may be sorted randomly into the descendent lineages. In this model, different segments of a genome may have different phylogenetic relationships. This process, which leads to conflicting evolutionary trees for different genes, is called incomplete lineage sorting (ILS; Box 1).

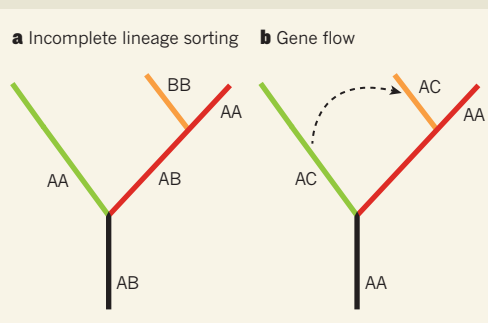
Previous molecular-genetic studies^{2,3} of humans, chimpanzees and gorillas show that the three lineages separated over a relatively short period of time, creating the opportunity for ILS. Scally and colleagues' whole-genome study verifies and substantially extends this analysis. They found that for 70% of the genomes of the three species, the chimpanzee sequences are more similar to the corresponding sequences in humans than to those in the gorilla, as expected. But for the remaining 30% of the genome, gorilla sequences share closer similarity with either human or chimpanzee sequences than these two share with each other. The authors also observed these ILS patterns reflected in relative levels of gene expression in the three species.

Other factors, such as gene flow between species after their initial divergence (Box 1), may also have contributed to these surprising relationships. There is evidence⁴ for such gene flow between Neanderthals and the lineage that ultimately produced modern humans, and between morphologically differentiated non-human primate groups that

BOX 1

How incongruities in phylogenetic trees can arise

Scally and colleagues¹ found that in 30% of the western-lowland-gorilla genome, the DNA sequences are more similar to the corresponding sequences from the human or chimpanzee genomes than the sequences of these two species are to each other — although humans and chimpanzees are expected to have shared a more recent common ancestor with each other than either does with gorillas. Such inconsistencies between evolutionary relationships can result from various processes. **a**, One possible mechanism is incomplete lineage sorting. In this process, an ancestral species (black) divides into two descendent genetic lineages (green and red) and, soon after, one of those descendent lineages divides again (red and



orange). The red and orange lineages are expected to be more genetically similar to one another than either is to the green lineage. However, if the ancestral species contained a gene with two alternative sequence variants

(AB), either or both variants may be transmitted into the descendants. Over time, the descendent species will lose one of the two gene variants. Here, incomplete lineage sorting has resulted in the red species being more genetically similar to the green species (both AA) than to the orange species (BB) at this particular gene. **b**, Gene flow is another mechanism by which relationships between specific DNA sequences can fail to match the larger relationships between species.

In this case, a newly evolved gene variant (C) is transferred from one genetic lineage to another by interbreeding that occurs after evolutionary separation has begun, but before complete genetic isolation is achieved. [R.A.G. & J.R.](#)

State University of New York at Buffalo,
Buffalo, New York 14214, USA.
e-mails: philgott@buffalo.edu; sachs@buffalo.edu

1. Chalfie, M. *Nature Rev. Mol. Cell. Biol.* **10**, 44–52 (2009).
2. Coste, B. *et al. Science* **330**, 55–60 (2010).
3. Coste, B. *et al. Nature* **483**, 176–181 (2012).
4. Kim, S. E., Coste, B., Chadha, A., Cook, B. & Patapoutian, A. *Nature* **483**, 209–212 (2012).
5. Guharay, F. & Sachs, F. *J. Physiol. (Lond.)* **352**, 685–701 (1984).
6. Bae, C., Sachs, F. & Gottlieb, P. A. *Biochemistry* **50**, 6295–6300 (2011).
7. Honoré, E., Patel, A. J., Chemin, J., Suchyna, T. & Sachs, F. *Proc. Natl Acad. Sci. USA* **103**, 6859–6864 (2006).
8. Kung, C., Martinac, B. & Sukharev, S. *Annu. Rev. Microbiol.* **64**, 313–329 (2010).

GENOMICS

Gorilla gorilla gorilla

The gorilla genome reveals that genetic similarities among humans and the apes are more complex than expected, and allows a fresh assessment of the evolutionary mechanisms that led to the primate species seen today. [SEE ARTICLE P.169](#)

RICHARD A. GIBBS & JEFFREY ROGERS

Humans and the apes are the living representatives of the superfamily Hominoidea, which also contains many extinct species. Deciphering the evolutionary relationships between these species is an essential step in our understanding of the biological richness of the planet, and of our own evolutionary history. The draft assembly of the whole-genome sequence of a female western lowland gorilla (*Gorilla gorilla gorilla*) named Kamilah (Fig. 1), presented by Scally *et al.*¹ on page 169 of this issue, provides insight into how a single hominid lineage separated into the extant human, chimpanzee and gorilla branches.

The authors' genome assembly, made possible by the advent of cheaper next-generation sequencing methods, is a much-anticipated

addition to more than five-year-old Sanger sequence data on the gorilla genome. The latest assembly, like other contemporary mammalian genome data sets (except for human and mouse), has some gaps and shortcomings. Yet it is revelatory. The standard view of the primate evolutionary tree is that chimpanzees and humans share a more recent common ancestor with each other than either shares with gorillas. Accordingly, the most closely related sequence for any human gene should be found in the chimpanzee. However, Scally and colleagues' demonstrate that, although this is true for most genes, large fractions of the ape genomes contradict this simple pattern.

Molecular phylogenetics uses comparative analyses of DNA sequences to determine relatedness among species. Usually, a single individual of each species is sequenced, so genetic diversity within a species is overlooked.

When the time between successive evolutionary branch points, or speciation events, is relatively long, the between-species genetic differences that accumulate after a single lineage divides into two descendent branches will stand out against the background of within-species variation. But if two or more evolutionary divergence events occur close together in time, the genetic variation present in the last common ancestor may be sorted randomly into the descendent lineages. In this model, different segments of a genome may have different phylogenetic relationships. This process, which leads to conflicting evolutionary trees for different genes, is called incomplete lineage sorting (ILS; Box 1).

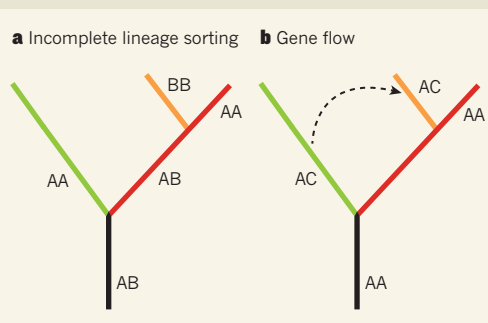
Previous molecular-genetic studies^{2,3} of humans, chimpanzees and gorillas show that the three lineages separated over a relatively short period of time, creating the opportunity for ILS. Scally and colleagues' whole-genome study verifies and substantially extends this analysis. They found that for 70% of the genomes of the three species, the chimpanzee sequences are more similar to the corresponding sequences in humans than to those in the gorilla, as expected. But for the remaining 30% of the genome, gorilla sequences share closer similarity with either human or chimpanzee sequences than these two share with each other. The authors also observed these ILS patterns reflected in relative levels of gene expression in the three species.

Other factors, such as gene flow between species after their initial divergence (Box 1), may also have contributed to these surprising relationships. There is evidence⁴ for such gene flow between Neanderthals and the lineage that ultimately produced modern humans, and between morphologically differentiated non-human primate groups that

BOX 1

How incongruities in phylogenetic trees can arise

Scally and colleagues¹ found that in 30% of the western-lowland-gorilla genome, the DNA sequences are more similar to the corresponding sequences from the human or chimpanzee genomes than the sequences of these two species are to each other — although humans and chimpanzees are expected to have shared a more recent common ancestor with each other than either does with gorillas. Such inconsistencies between evolutionary relationships can result from various processes. **a**, One possible mechanism is incomplete lineage sorting. In this process, an ancestral species (black) divides into two descendent genetic lineages (green and red) and, soon after, one of those descendent lineages divides again (red and



orange). The red and orange lineages are expected to be more genetically similar to one another than either is to the green lineage. However, if the ancestral species contained a gene with two alternative sequence variants

(AB), either or both variants may be transmitted into the descendants. Over time, the descendent species will lose one of the two gene variants. Here, incomplete lineage sorting has resulted in the red species being more genetically similar to the green species (both AA) than to the orange species (BB) at this particular gene. **b**, Gene flow is another mechanism by which relationships between specific DNA sequences can fail to match the larger relationships between species.

In this case, a newly evolved gene variant (C) is transferred from one genetic lineage to another by interbreeding that occurs after evolutionary separation has begun, but before complete genetic isolation is achieved. [R.A.G. & J.R.](#)



Figure 1 | Kamilah the gorilla. Scally *et al.*¹ report the whole-genome sequence of this western lowland gorilla.

are often considered separate species, such as macaques^{5,6} and baboons^{7,8}. Thus, the data presented by Scally and colleagues¹ support a growing challenge to the idea that primate species diversity always increases through a process in which one species divides rapidly and irreversibly into two genetically isolated daughter species.

New calculations to establish when the gorilla, chimpanzee and human lineages diverged are now possible. By combining data on DNA-sequence differences between species with estimates of DNA-mutation rates operative millions of years ago, Scally and colleagues calculate the human–chimpanzee and (human+chimpanzee)–gorilla branch points to be 3.7 million and 5.95 million years ago, respectively. These dates conflict with some fossil evidence, but more complex calculations that allow for mutation rates that change with time — a reasonable possibility owing to changes in the average length of generations and other parameters — may resolve this problem.

An overarching question is how these three divergent lineages each acquired their unique morphological and behavioural characteristics — including how one lineage became human. The new gorilla data suggest that a large proportion of the human genome was under positive selection pressure during the period of initial isolation from our close relatives, indicating that natural selection drove evolutionary innovation and change. Scally and colleagues' study adds to our ability to reconstruct those selective forces that acted on each lineage.

The report¹ also provides information on genetic diversity within the *Gorilla* genus. The authors compared Kamilah's genome with sequence data from two other western lowland gorillas and an eastern lowland gorilla (*Gorilla beringei graueri*), and found substantial genetic differentiation within the genus, consistent

with previous data⁹. The genus is currently classified into four subspecies, two for each of the two species, but the variation identified by Scally *et al.* suggests further analysis of genetic diversity in gorillas is warranted.

Gorillas have inspired awe and scientific interest throughout history. The sequencing of the gorilla genome adds valuable information to our understanding of these remarkable animals and our evolutionary relationship to them — revealing a closer connection between our genome and that of the gorilla than was previously appreciated. Sadly, the future of gorillas is uncertain, with all populations under severe human-induced pressure, and the mountain gorillas especially close to extinction. Their fate is inextricably linked to our choices, and these extraordinary primates deserve our utmost efforts at conservation. To quote from American author Daniel Quinn's novel *Ishmael*, “with man gone, will there be hope for

gorilla?” or, even more soberingly, “with gorilla gone, will there be hope for man?” ■

Richard A. Gibbs and Jeffrey Rogers are at the Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA.
e-mails: agibbs@bcm.edu; jr13@bcm.edu

1. Scally, A. *et al.* *Nature* **483**, 169–175 (2012).
2. Rogers, J. *Am. J. Phys. Anthropol.* **94**, 81–88 (1994).
3. Ruvolo, M. *Mol. Biol. Evol.* **14**, 248–265 (1997).
4. Green, R. E. *et al.* *Science* **328**, 710–722 (2010).
5. Tosi, A. J., Morales, J. C. & Melnick, D. J. *Int. J. Primatol.* **23**, 161–178 (2002).
6. Osada, N. *et al.* *Mol. Ecol.* **19**, 2884–2895 (2010).
7. Alberts, S. C. & Altmann, J. *Am. J. Primatol.* **53**, 139–154 (2001).
8. Jolly, C. J., Burrell, A. S., Phillips-Conroy, J. E., Bergcey, C. & Rogers, J. *Am. J. Primatol.* **73**, 291–303 (2011).
9. Thalmann, O., Fischer, A., Lankester, F., Pääbo, S. & Vigilant, L. *Mol. Biol. Evol.* **24**, 146–158 (2007).

GEOCHEMISTRY

A rusty carbon sink

The finding that reactive iron species may have a role in stabilizing organic matter in ocean sediments underlines the tight coupling between the biogeochemical cycles of carbon and iron. [SEE LETTER P.198](#)

TIM I. EGLINTON

The sequestration of organic matter in marine sediments is a primary mechanism by which carbon is removed from the atmosphere and stored over geological periods of time. Only a tiny fraction of organic matter produced by the biosphere arrives at, and is eventually buried in, the sediment bed; the vast majority is recycled back to carbon dioxide. But how and why does any organic carbon remain to be incorporated into the sedimentary record? On page 198 of this issue, Lalonde and co-workers¹ report findings that might help to solve this puzzle. They reveal that the intimate association of organic matter with iron species in sediments — including some of the iron compounds found in rust — may exert a strong influence on carbon burial.

Despite decades of research, our understanding of the mechanisms by which organic matter is stabilized and sequestered in aquatic sediments remains far from complete², hindering our ability to develop robust theories of carbon burial. In recent years, a model has emerged proposing that the intrinsic reactivity of the organic matter supplied to soils and sediments is less important than its distribution in the sedimentary matrix. This, along with a few other factors³, controls its accessibility to the organisms that degrade it into simple organic compounds, and, ultimately, to carbon dioxide.

In soils, the formation of complexes of iron salts and organic matter has been identified as one mechanism for stabilizing organic carbon. But despite evidence of a close coupling of organic carbon and iron in marine sediments, this relationship has not previously been explored as a mechanism for preserving organic matter in aquatic environments.

Lalonde *et al.*¹ used a simple chemical procedure⁴ to release reactive forms of iron from sediment samples and to quantify the amount of organic carbon mobilized during this treatment. Their findings suggest that, on average, more than 20% of the organic carbon in aquatic sediments from a wide range of depositional environments — which vary in salinity, proximity to land, water depth, organic carbon content and oxygen availability — is associated with reactive iron. The authors also measured similar proportions of iron-bound organic carbon in older, deeper sediments, suggesting that such organic carbon exhibits marked stability and prolonged resistance to microbial attack.

Because the available surface area of iron minerals in sediments is insufficient to accommodate the observed organic-carbon loadings, Lalonde *et al.* conclude that adsorption of organic matter to iron-containing matter is insufficient to explain their observations. Instead, they propose that iron salts must have co-precipitated with organic carbon or that iron–organic-carbon complexes must



Figure 1 | Kamilah the gorilla. Scally *et al.*¹ report the whole-genome sequence of this western lowland gorilla.

are often considered separate species, such as macaques^{5,6} and baboons^{7,8}. Thus, the data presented by Scally and colleagues¹ support a growing challenge to the idea that primate species diversity always increases through a process in which one species divides rapidly and irreversibly into two genetically isolated daughter species.

New calculations to establish when the gorilla, chimpanzee and human lineages diverged are now possible. By combining data on DNA-sequence differences between species with estimates of DNA-mutation rates operative millions of years ago, Scally and colleagues calculate the human–chimpanzee and (human+chimpanzee)–gorilla branch points to be 3.7 million and 5.95 million years ago, respectively. These dates conflict with some fossil evidence, but more complex calculations that allow for mutation rates that change with time — a reasonable possibility owing to changes in the average length of generations and other parameters — may resolve this problem.

An overarching question is how these three divergent lineages each acquired their unique morphological and behavioural characteristics — including how one lineage became human. The new gorilla data suggest that a large proportion of the human genome was under positive selection pressure during the period of initial isolation from our close relatives, indicating that natural selection drove evolutionary innovation and change. Scally and colleagues' study adds to our ability to reconstruct those selective forces that acted on each lineage.

The report¹ also provides information on genetic diversity within the *Gorilla* genus. The authors compared Kamilah's genome with sequence data from two other western lowland gorillas and an eastern lowland gorilla (*Gorilla beringei graueri*), and found substantial genetic differentiation within the genus, consistent

with previous data⁹. The genus is currently classified into four subspecies, two for each of the two species, but the variation identified by Scally *et al.* suggests further analysis of genetic diversity in gorillas is warranted.

Gorillas have inspired awe and scientific interest throughout history. The sequencing of the gorilla genome adds valuable information to our understanding of these remarkable animals and our evolutionary relationship to them — revealing a closer connection between our genome and that of the gorilla than was previously appreciated. Sadly, the future of gorillas is uncertain, with all populations under severe human-induced pressure, and the mountain gorillas especially close to extinction. Their fate is inextricably linked to our choices, and these extraordinary primates deserve our utmost efforts at conservation. To quote from American author Daniel Quinn's novel *Ishmael*, “with man gone, will there be hope for

gorilla?” or, even more soberingly, “with gorilla gone, will there be hope for man?” ■

Richard A. Gibbs and Jeffrey Rogers are at the Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA.
e-mails: agibbs@bcm.edu; jr13@bcm.edu

1. Scally, A. *et al.* *Nature* **483**, 169–175 (2012).
2. Rogers, J. *Am. J. Phys. Anthropol.* **94**, 81–88 (1994).
3. Ruvolo, M. *Mol. Biol. Evol.* **14**, 248–265 (1997).
4. Green, R. E. *et al.* *Science* **328**, 710–722 (2010).
5. Tosi, A. J., Morales, J. C. & Melnick, D. J. *Int. J. Primatol.* **23**, 161–178 (2002).
6. Osada, N. *et al.* *Mol. Ecol.* **19**, 2884–2895 (2010).
7. Alberts, S. C. & Altmann, J. *Am. J. Primatol.* **53**, 139–154 (2001).
8. Jolly, C. J., Burrell, A. S., Phillips-Conroy, J. E., Bergcey, C. & Rogers, J. *Am. J. Primatol.* **73**, 291–303 (2011).
9. Thalmann, O., Fischer, A., Lankester, F., Pääbo, S. & Vigilant, L. *Mol. Biol. Evol.* **24**, 146–158 (2007).

GEOCHEMISTRY

A rusty carbon sink

The finding that reactive iron species may have a role in stabilizing organic matter in ocean sediments underlines the tight coupling between the biogeochemical cycles of carbon and iron. SEE LETTER P.198

TIM I. EGLINTON

The sequestration of organic matter in marine sediments is a primary mechanism by which carbon is removed from the atmosphere and stored over geological periods of time. Only a tiny fraction of organic matter produced by the biosphere arrives at, and is eventually buried in, the sediment bed; the vast majority is recycled back to carbon dioxide. But how and why does any organic carbon remain to be incorporated into the sedimentary record? On page 198 of this issue, Lalonde and co-workers¹ report findings that might help to solve this puzzle. They reveal that the intimate association of organic matter with iron species in sediments — including some of the iron compounds found in rust — may exert a strong influence on carbon burial.

Despite decades of research, our understanding of the mechanisms by which organic matter is stabilized and sequestered in aquatic sediments remains far from complete², hindering our ability to develop robust theories of carbon burial. In recent years, a model has emerged proposing that the intrinsic reactivity of the organic matter supplied to soils and sediments is less important than its distribution in the sedimentary matrix. This, along with a few other factors³, controls its accessibility to the organisms that degrade it into simple organic compounds, and, ultimately, to carbon dioxide.

In soils, the formation of complexes of iron salts and organic matter has been identified as one mechanism for stabilizing organic carbon. But despite evidence of a close coupling of organic carbon and iron in marine sediments, this relationship has not previously been explored as a mechanism for preserving organic matter in aquatic environments.

Lalonde *et al.*¹ used a simple chemical procedure⁴ to release reactive forms of iron from sediment samples and to quantify the amount of organic carbon mobilized during this treatment. Their findings suggest that, on average, more than 20% of the organic carbon in aquatic sediments from a wide range of depositional environments — which vary in salinity, proximity to land, water depth, organic carbon content and oxygen availability — is associated with reactive iron. The authors also measured similar proportions of iron-bound organic carbon in older, deeper sediments, suggesting that such organic carbon exhibits marked stability and prolonged resistance to microbial attack.

Because the available surface area of iron minerals in sediments is insufficient to accommodate the observed organic-carbon loadings, Lalonde *et al.* conclude that adsorption of organic matter to iron-containing matter is insufficient to explain their observations. Instead, they propose that iron salts must have co-precipitated with organic carbon or that iron–organic-carbon complexes must

have formed, or both, in diagenetic processes — those involving iron species made available near the sediment surface as a result of biogeochemical or physical changes that occurred shortly after sedimentation.

More specifically, the authors envisage the formation of nanometre-scale domains consisting of macromolecular complexes formed from organic carbon and iron oxide(hydroxide) minerals, and posit that these domains are sufficiently stable to persist for many thousands of years. Their results indicate that the carbon loading of these complexes may decrease during long-term oxygen exposure, but the overall apparent stability of the complexes, particularly under anoxic conditions, implies that they can provide a mechanism for the long-term sequestration of organic carbon.

Further work is needed to confirm that Lalonde and colleagues' chemical analysis specifically detects only the organic matter bound in iron complexes, given that the method has previously been applied only to studies of soil matrices⁴. But assuming the observations are robust, the authors' findings are impressive from a variety of perspectives. First, they represent a notable advance in our understanding of how organic matter can be preserved in aquatic sediments, one that draws parallels with similar organic-carbon stabilization processes that are prevalent in soils⁵.

Second, the chemical characteristics of the organic matter associated with the iron complexes suggest that it is enriched in biochemical components that are typically considered to be highly degradable. This observation may help to explain the apparent paradox of how intrinsically labile biomolecules can be preserved in the sedimentary record.

Third, if the overall fraction of organic carbon bound in iron complexes is similar to that observed for the samples measured in this study, then such complexes represent a globally important contributor to the long-term sedimentary sink of organic carbon. Taken together, the findings underline the tight coupling between the biogeochemical cycles of carbon and iron.

However, the authors' study¹ leaves many open questions. For example, the precise mechanism by which iron stabilizes organic matter remains unknown, as does the nature and evolution of iron-organic-carbon complexes during early diagenesis and longer-term burial. The selectivity of the stabilization processes for different types of organic compound involved in the formation of iron-organic-carbon complexes, and the impact of this on the legacy of past biological activity preserved as organic signatures in the rock record, also deserve further attention. We can only speculate on how these processes have influenced organic-carbon burial during periods when oceanic conditions were significantly different from today's — for example, because of changes in sea level or concentrations of

dissolved oxygen. Overall, however, the recognition of a 'rusty sink' for organic carbon in ocean sea-floor sediments represents a major advance in our understanding of this enigmatic component of the carbon cycle. ■

Tim I. Eglinton is at the *Geologisches Institut, ETH Zürich, 8092 Zürich, Switzerland.*

AGEING

Sorting out the sirtuins

Debates over the role of sirtuin proteins in ageing are maturing into functional assessments of the individual proteins. It seems that overexpression of a specific sirtuin can extend lifespan in male mice. [SEE LETTER P.218](#)

DAVID B. LOMBARD & RICHARD A. MILLER

Abraham Lincoln once said that God must have loved the common people because he made so many of them. Nature must feel the same way about the sirtuins, a large family of proteins that achieved celebrity status when one member was found to increase lifespan in yeast¹. But are the mammalian sirtuins the rock stars of an ensemble of anti-ageing proteins, or merely members of the entourage? The original model, proposed in about 2005, that sirtuins have broadly evolutionarily conserved roles in promoting longevity per se is now being refined through more detailed functional investigations of each sirtuin². On page 218 of this issue, Kanfi *et al.*³ follow this trend by reporting that overexpression of a sirtuin called SIRT6 leads to a modest extension of lifespan in male, but not female, mice^{*}.

Does an extension of lifespan imply an effect on ageing? Not necessarily: interventions unrelated to ageing, such as giving insulin to a person with type I diabetes, can increase mean and maximal lifespan. The lifespan extension observed by Kanfi and colleagues in SIRT6-overexpressing male mice could be explained, at least partially, by SIRT6 acting as a tumour suppressor. Because male mice have a higher incidence of spontaneous cancer than female mice (incidences of 81% and 50%, respectively, were observed in this study), an anticancer protein (perhaps SIRT6?) would have a larger effect on lifespan in males than in females.

Proving that a lifespan-increasing intervention indeed acts by delaying ageing processes is not a simple matter. For example, acceptance of the idea that lifespan extension by caloric restriction (a diet with reduced calorie intake) reflects a genuine deceleration of ageing emerged gradually from evidence⁴ that restriction slows age-related changes in the

e-mail: timothy.eglinton@erdw.ethz.ch

1. Lalonde, K., Mucci, A., Ouellet, A. & Gélinas, Y. *Nature* **483**, 198–200 (2012).
2. Hedges, J. I. & Keil, R. G. *Mar. Chem.* **49**, 81–115 (1995).
3. Schmidt, M. W. I. *et al. Nature* **478**, 49–56 (2011).
4. Mehra, O. P. & Jackson, M. L. *Clays Clay Miner.* **7**, 317–327 (1958).
5. Kaiser, K. & Guggenberger, G. *Org. Geochem.* **31**, 711–725 (2000).

properties of proliferative and non-proliferative cells in many tissues, and does so in multiple organ systems. Similar cases are being constructed by researchers proposing that dwarf mice could act as models for slowed ageing⁵.

Reports of lifespan increases in mutant or drug-treated mice, particularly studies in which the observed effects are modest, often prove difficult for other laboratories to repeat. This is presumably due to subtle but crucial variations in the animals' diet or genetic background, or in husbandry practices⁶. Moreover, the preference for publication of positive over negative findings inevitably inserts a smattering of false positive results into the literature, and these can be identified only by attempts to replicate experiments. One strength of Kanfi and colleagues' paper³ is that SIRT6 overexpression increased male lifespan in each of two groups of mice, which were derived from two different founder animals. However, the test for maximal lifespan — usually taken as stronger evidence than an effect on median longevity alone — reached statistical significance in only one of the two mouse groups. If the longevity effect seen by the authors proves robust, determining whether SIRT6 overexpression does indeed slow ageing will still require follow-up studies analysing a wide range of age-sensitive endpoints.

In their article, Kanfi and colleagues include some observations hinting at potential mechanisms by which SIRT6 overexpression might affect the lifespan of male mice. Compared with their normal counterparts, SIRT6-overexpressing males had modestly reduced serum levels of the hormone IGF-1, and the signalling activity of IGF-1 receptors was weaker in perigonadal fat tissue in males but not in females. Previous reports have found that SIRT6 attenuates intracellular signalling initiated by IGF-1 and insulin⁷. Furthermore, dramatic deficits in IGF-1 and/or growth hormone (GH, which stimulates IGF-1 secretion)

^{*}This article and the paper³ under discussion were published online on 22 February 2012.

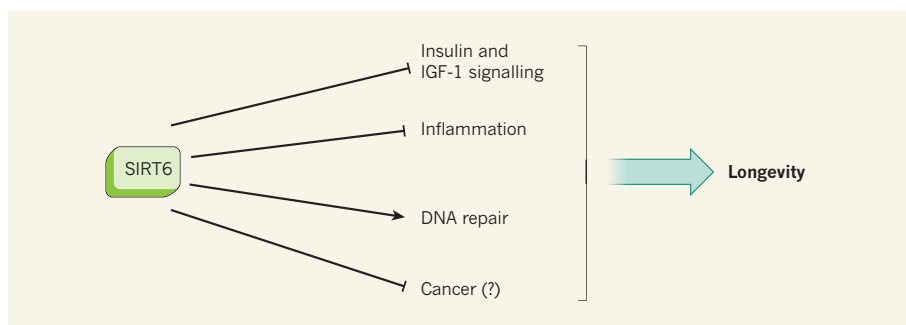


Figure 1 | Potential mechanisms of action of SIRT6 on longevity. Several reports^{7,10–12} have demonstrated effects of the sirtuin protein SIRT6 on the activities of the hormones insulin and IGF-1, as well as on inflammation and DNA repair. These effects, together with a possible delay in cancer progression, could contribute to the increased lifespan in SIRT6-overexpressing male mice reported by Kanfi and colleagues³.

lead to slower ageing and increased lifespan in at least four varieties of mutant mouse⁶. And mutations in the gene encoding the GH receptor in humans are associated with strong protection against diabetes and cancer⁸. So, it is plausible that SIRT6 overexpression in mice might work through blunting of the GH/IGF-1 pathway. Evidence⁹ that rat longevity can be augmented by surgical removal of intra-abdominal — but not subcutaneous — fat has begun to focus attention on metabolic and hormonal effects on specific fat depots as potential levers for pharmacological control of ageing.

It is noteworthy that the effects of SIRT6 overexpression reported by Kanfi *et al.*³ are seen only in male mice. Previous results⁶, by contrast, indicate that mutations in components of the GH/IGF-1 pathway usually have greater effects on longevity in female mice. This apparent discrepancy might be explained by differences between the mice in terms of underlying disease proclivities, levels of sex-specific hormones, inter-animal conflict or fat-tissue biology, leading to gender-specific responses to mutations, drugs and nutritional interventions. Working out the basis for these sex-specific interactions should provide clues to the mechanisms involved in these anti-ageing manipulations, and perhaps even help to answer the vexing question of why women tend to live longer than men.

SIRT6 has other roles that could foster longer lifespan (Fig. 1). It promotes chromosomal stability by several mechanisms, and above-normal SIRT6 expression increases the efficiency of DNA repair¹⁰. SIRT6 also reduces the expression of genes regulated by the NF- κ B and HIF-1 α proteins, which have roles in inflammation, cancer and, potentially, longevity^{11,12}. It will be of interest to assess these aspects of SIRT6's function in mice overexpressing the protein, and to test more definitively whether they contribute to protection against cancer and promotion of longevity.

The recent spate of activity in sirtuin research, now supplemented by the present work, supports the case

for placing the sirtuins on the front line of ageing research, sitting cheek by jowl with other promising contestants, such as the proteins TOR, FoxO, AMPK, NRF2 and ATF4. To paraphrase Winston Churchill, the discoveries of Kanfi *et al.* do not by any means represent the end of sirtuin research, nor even the beginning

of the end. But they are, perhaps, the end of the beginning. ■

David B. Lombard and Richard A. Miller are in the Department of Pathology and the Geriatrics Center, University of Michigan School of Medicine, Ann Arbor, Michigan 48100-2200, USA.
e-mails: davidlom@umich.edu;
millerr@umich.edu

1. Kaeberlein, M., McVey, M. & Guarente, L. *Genes Dev.* **13**, 2570–2580 (1999).
2. Finkel, T., Deng, C.-X. & Mostoslavsky, R. *Nature* **460**, 587–590 (2009).
3. Kanfi, Y. *et al.* *Nature* **483**, 218–221 (2012).
4. Weindruch, R. & Sohal, R. S. N. *Engl. J. Med.* **337**, 986–994 (1997).
5. Bartke, A. *Phil. Trans. R. Soc. Lond. B* **366**, 28–34 (2011).
6. Ladiges, W. *et al.* *Ageing Cell* **8**, 346–352 (2009).
7. Xiao, C. *et al.* *J. Biol. Chem.* **285**, 36776–36784 (2010).
8. Guevara-Aguirre, J. *et al.* *Sci. Transl. Med.* **3**, 70ra13 (2011).
9. Muzumdar, R. *et al.* *Ageing Cell* **7**, 438–440 (2008).
10. Mao, Z. *et al.* *Science* **332**, 1443–1446 (2011).
11. Kawahara, T. L. A. *et al.* *Cell* **136**, 62–74 (2009).
12. Zhong, L. *et al.* *Cell* **140**, 280–293 (2010).

MATERIALS SCIENCE

Continuity through dispersity

By making polymers whose central blocks have a range of lengths, materials have been prepared that contain separate, intermeshed domains extending throughout the material — a highly desirable structure.

RICHARD A. REGISTER

Ideas about nanometre-scale self-assembly — the phenomenon in which nanoscale objects organize themselves into arrays — are commonly underpinned by two assumptions. The first is that the higher the degree of regularity the array has, the better. The second is that obtaining a high degree of regularity requires the assembling objects to be as monodisperse (identical in shape and size) as possible. But these assumptions have recently come under intense scrutiny in the case of self-assembling block copolymers. The molecules in these materials consist of long runs (blocks) made from different monomer types, which can separate into distinct nanoscale domains. Writing in the *Journal of the American Chemical Society*, Mahanthappa and colleagues¹ report that molecules of 'ABA' triblock copolymers (in which A and B represent different blocks) robustly generate desirable self-assembled structures containing separate A and B domains when the middle blocks have a broad range of lengths.

For many applications, it is preferable for

the A and B domains of a block copolymer to be bicontinuous — that is, to interweave and extend throughout the material. For example, in photovoltaic films used in solar cells, excitons (electron–hole pairs, where holes are quasiparticles formed by the absence of electrons) generated by photon absorption could rapidly diffuse to interfaces between bicontinuous A and B domains, thus allowing the charges to separate, find their way out of the film and so generate a current^{2,3}. Or consider the separator membranes in batteries, which divide the anode from the cathode but provide a medium through which ions can pass. A bicontinuous material could provide both a soft, rubbery domain that would facilitate ion diffusion, and a hard, glassy domain that would maintain the membrane's mechanical integrity and stiffness⁴.

Unfortunately, the regular bicontinuous structures (such as the gyroid phase; Fig. 1a) that form in conventional, near-monodisperse copolymers consisting of only two types of block occupy frustratingly narrow slices of 'composition space' — if the fraction of material occupied by A blocks is expressed

► **NATURE.COM**
For more on
sirtuins, see:
go.nature.com/8shahko

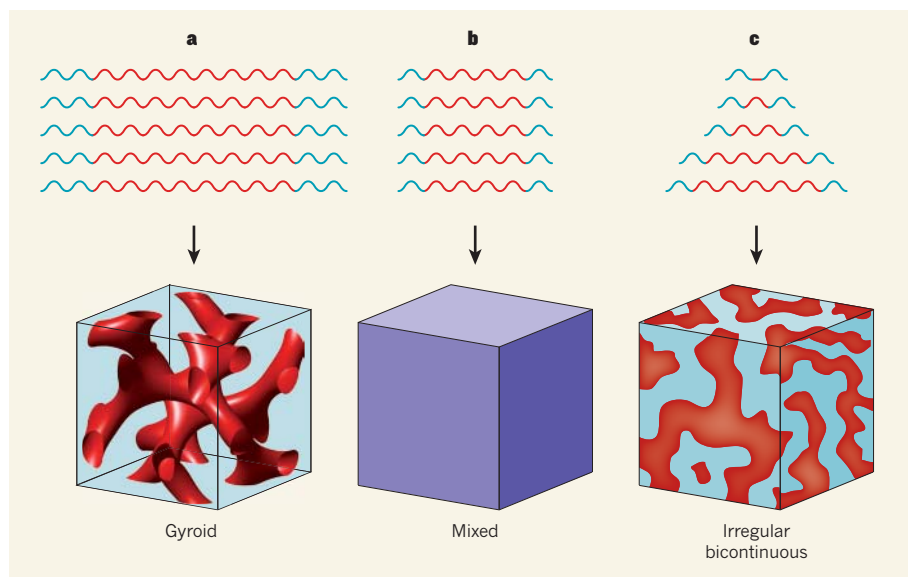


Figure 1 | Self-assembly of copolymers. ABA triblock copolymer chains consist of two A blocks (blue) separated by a B block (red), where the A and B blocks are formed from different monomers. **a**, In this case, all blocks of the same type are the same length, and the blocks are long enough to be incompatible — the different types don't mix. At certain ratios of the lengths of A and B, the molecules instead self-assemble into a regular bicontinuous structure (known as a gyroid) containing separate, intermeshed domains of A and B that span the whole material. **b**, Here, all blocks of the same type are again the same length, but the blocks are short enough to mix with each other. No separate A or B domains form. **c**, Mahanthappa and colleagues¹ report that when the B blocks of the polymer have a range of lengths, an irregular bicontinuous structure forms robustly — even though the A blocks are the same length as in **b**, and the average length of the B blocks is shorter than in **b**. (Image in **a** courtesy Daniel C. Fredrickson; image in **c** courtesy Andrew L. Schmitt.)

as a percentage, then the range of values for which regular bicontinuous domains can form is only about 3%. But Mahanthappa and colleagues¹ report that an irregular bicontinuous structure⁵ can be obtained over a broader composition window of about 10%, simply by introducing substantial polydispersity (a range of polymer chain lengths) into an ABA triblock copolymer. Previous studies have shown that AB diblocks⁶ and BAB triblocks⁷ that have polydisperse B blocks do not form such a bicontinuous structure, so it seems that the key to success is for the centre block to be polydisperse.

Mahanthappa and co-workers found that the irregular bicontinuous structure is not the only morphology that their ABA triblock copolymers can adopt. At other compositions (A:B ratios), they form well-ordered lamellae or cylinders, again despite the chain-length variation. But the bicontinuous structure is especially stable. To explain what this means in more detail, we need to consider the factors that affect the formation of such domain structures.

In any block copolymer, when the blocks become too short, they no longer separate into domains. Instead, they simply mix (Fig. 1b). The driving force for block segregation is parameterized as the product χN , where χ is the Flory interaction parameter (a measure of the strength of repulsive interactions between the monomers that make up A and B blocks, independent of block length) and N is the total

degree of polymerization of the block copolymer (a measure of the copolymer's length). For monodisperse ABA triblock copolymers, the minimum value of χN for domain formation⁸ is close to 18. But in Mahanthappa and colleagues' triblocks, which have near-monodisperse endblocks and a polydisperse mid-block, the minimum χN value required to form the disordered bicontinuous structure is only about 5, an enormous reduction. In practice, this means that even polymers that have fairly low molar masses — an average of 12 kilograms per mole for the authors' ABA triblocks, in which A was polystyrene and B was polybutadiene — are sufficient to generate the sought-after bicontinuous domain structure.

Near-monodisperse block copolymers are conventionally synthesized using one of several controlled polymerization mechanisms. In these syntheses, the block sequence is dictated simply by the order in which different monomers are charged into the reaction mixture; each monomer is allowed to polymerize fully, or excess unreacted monomer is removed, before the next monomer is added. If a particular mechanism cannot polymerize a given monomer to sufficiently high average molar mass and/or if it cannot yield a product that has a sufficiently narrow distribution of molar masses, then it is generally discarded as a route to the target polymer. But Mahanthappa and co-workers' findings show that neither a narrow distribution nor a particularly high molar

mass is required to make bicontinuous structures. Instead, all that seems necessary is that the polydisperse B blocks should be tethered at a minimum of two points, as they are in ABA triblock copolymers.

Less clear at present is whether asymmetric polydispersity is also required — that is, whether the lengths of the A blocks must be narrowly distributed, as they are in Mahanthappa and colleagues' polymers. A narrow distribution of A-block lengths causes asymmetric block mixing, which in turn affects the compositions over which different phases (cylinders, bicontinuous structures and lamellae) form. For example, if some of the B blocks in a sample of an ABA triblock copolymer are short, such as those depicted at the top of Fig. 1c, then they will readily dissolve in the A domains. But if all the A blocks are essentially the same length, then relatively few of them will dissolve in B domains. As a consequence, the fraction of the volume of the material that consists of B-rich domains will differ substantially from what one would expect assuming complete separation of A and B into pure domains.

It remains to be seen how the mixing of A and B blocks in Mahanthappa and colleagues' polydisperse polymers affects the physical properties of the materials — such as the glass transition temperature, which is important both for diffusion of small molecules or ions through rubbery domains and for the mechanical integrity of glassy domains. Another question is whether both domains are fully continuous (sample-spanning) across the entire composition range for which the irregular bicontinuous structure forms. Follow-up studies investigating these issues will surely come quickly. In the meantime, we should add mid-block polydispersity to the toolbox of macromolecular architecture variations⁹ that allows the synthesis of polymers that have predictable self-assembled structures and useful properties. ■

Richard A. Register is in the Department of Chemical and Biological Engineering, Princeton University, Princeton, New Jersey 08544, USA.

e-mail: register@princeton.edu

- Widin, J. M., Schmitt, A. K., Schmitt, A. L., Im, K. H. & Mahanthappa, M. K. *J. Am. Chem. Soc.* <http://dx.doi.org/10.1021/ja210548e> (2012).
- Crossland, E. J. W. *et al. Nano Lett.* **9**, 2807–2812 (2009).
- Ho, V. *et al. J. Am. Chem. Soc.* **133**, 9270–9273 (2011).
- Singh, M. *et al. Macromolecules* **40**, 4578–4585 (2007).
- Widin, J. M., Schmitt, A. K., Im, K., Schmitt, A. L. & Mahanthappa, M. K. *Macromolecules* **43**, 7913–7915 (2010).
- Lynd, N. A., Meuler, A. J. & Hillmyer, M. A. *Prog. Polym. Sci.* **33**, 875–893 (2008).
- Ruzette, A.-V. *et al. Macromolecules* **39**, 5804–5814 (2006).
- Mayes, A. M. & Olvera de la Cruz, M. *J. Chem. Phys.* **91**, 7228–7235 (1989).
- Matyjaszewski, K. *Science* **333**, 1104–1105 (2011).

The role of *Drosophila* Piezo in mechanical nociception

Sung Eun Kim¹, Bertrand Coste¹, Abhishek Chadha¹, Boaz Cook¹ & Ardem Patapoutian^{1,2}

Transduction of mechanical stimuli by receptor cells is essential for senses such as hearing, touch and pain^{1–4}. Ion channels have a role in neuronal mechanotransduction in invertebrates¹; however, functional conservation of these ion channels in mammalian mechanotransduction is not observed. For example, no mechanoreceptor potential C (NOMPC), a member of transient receptor potential (TRP) ion channel family, acts as a mechanotransducer in *Drosophila melanogaster*⁵ and *Caenorhabditis elegans*^{6,7}; however, it has no orthologues in mammals. Degenerin/epithelial sodium channel (DEG/ENaC) family members are mechanotransducers in *C. elegans*⁸ and potentially in *D. melanogaster*⁹; however, a direct role of its mammalian homologues in sensing mechanical force has not been shown. Recently, Piezo1 (also known as Fam38a) and Piezo2 (also known as Fam38b) were identified as components of mechanically activated channels in mammals¹⁰. The Piezo family are evolutionarily conserved transmembrane proteins. It is unknown whether they function in mechanical sensing *in vivo* and, if they do, which mechanosensory modalities they mediate. Here we study the physiological role of the single Piezo member in *D. melanogaster* (*Dmpiezo*; also known as CG8486). *Dmpiezo* expression in human cells induces mechanically activated currents, similar to its mammalian counterparts¹¹. Behavioural responses to noxious mechanical stimuli were severely reduced in *Dmpiezo* knockout larvae, whereas responses to another noxious stimulus or touch were not affected. Knocking down *Dmpiezo* in sensory neurons that mediate nociception and express the DEG/ENaC ion channel *pickpocket* (*ppk*) was sufficient to impair responses to noxious mechanical stimuli. Furthermore, expression of *Dmpiezo* in these same neurons rescued the phenotype of the constitutive *Dmpiezo* knockout larvae. Accordingly, electrophysiological recordings from *ppk*-positive neurons revealed a *Dmpiezo*-dependent, mechanically activated current. Finally, we found that *Dmpiezo* and *ppk* function in parallel pathways in *ppk*-positive cells, and that mechanical nociception is abolished in the absence of both channels. These data demonstrate the physiological relevance of the Piezo family in mechanotransduction *in vivo*, supporting a role of Piezo proteins in mechanosensory nociception.

D. melanogaster is widely used to study mechanotransduction, and genetic studies have identified several ion channels that are essential for mechanosensation^{5,9,12–14}. However, none of the identified proteins have been shown to be activated by mechanical force when expressed in heterologous systems. Because expression of mouse Piezo proteins in a variety of mammalian cells induces mechanically activated currents¹⁰, we investigated whether the *Drosophila* counterpart is also sufficient to induce mechanosensitivity. Similar to its mammalian counterparts, the *Dmpiezo* gene is predicted to consist of a large number of transmembrane domains (39; Supplementary Fig. 1). Although fly and mammalian *piezo* genes do not exhibit extensive sequence conservation (24% identity), expression of *Dmpiezo* in cultured human cells induced mechanically activated cationic currents, suggesting a role of *Dmpiezo* in mechanotransduction¹¹.

To characterize *Dmpiezo* expression in flies we used a fusion between the *Dmpiezo* enhancer/promoter region and GAL4 (*DmpiezoP*–GAL4). Four independent *DmpiezoP*–GAL4 transgenic insertions were examined to avoid insertional effects on GAL4 expression. We used green fluorescent protein (GFP) regulated by upstream activating sequence elements (UAS) (UAS-GFP) for labelling cells, except for arborized neurons that were optimally visualized using the membrane-targeted UAS-CD8::GFP. We found fluorescent labelling induced by *Dmpiezo* enhancer/promoter region in all types of sensory neurons and several non-neuronal tissues in both adults and larvae (Supplementary Fig. 2). This diverse pattern of *Dmpiezo* expression observed in *Drosophila* is in accord with the expression of Piezo1 and Piezo2 in mice¹⁰.

We created *Dmpiezo* knockout flies in which all 31 coding exons were deleted using genomic recombination¹⁵ (Fig. 1a, see Supplementary Fig. 3 for details). The knockout flies were viable, fertile and did not show a lack of coordination or a defect in bristle mechanoreceptor potential (Supplementary Fig. 4). We studied whether *Dmpiezo* knockout larvae have mechanical nociception deficits by using a mechanically induced escape behaviour assay^{9,14,16}. Stimulation with von Frey filaments that ranged from 2–60 milliNewton (mN) demonstrated that *Dmpiezo* knockout larvae have a severe response deficit over a wide range (Fig. 1b). Repeated stimulations of the same larvae resulted in comparable responsiveness in both wild-type and *Dmpiezo* knockout, indicating that the stimuli did not induce considerable damage to the sensory system (Fig. 1c, d). A 153 ± 11.0 mN filament elicited responses only to the first of three stimulations in wild-type larvae, arguing that this amount of force is damaging (data not shown). For further experiments, we chose to stimulate the larvae using a 45 mN filament, which has been used in a previous study¹⁴, and elicits a substantial response in both wild-type and *Dmpiezo* mutant larvae. Thirty four ± 4.4% of *Dmpiezo* knockout larvae showed a response to 45 mN filament stimulation, compared to over 80% of wild-type or heterozygote larvae (Fig. 1e). As a control for the genetic background, we used larvae that carry the *Dmpiezo* knockout allele on one chromosome and a deficiency in which the entire *Dmpiezo* genomic region is deleted on the homologous chromosome. The defect in the trans-heterozygote larvae was similar to the knockout homozygote phenotype (51 ± 3.9%, *P* = 0.091). In contrast, *Dmpiezo* knockout larvae were indistinguishable from wild type in an assay for responses to high temperature, a different noxious stimulus that elicits the same escape response¹⁴ (Fig. 1f). Therefore, *Dmpiezo* knockout larvae retain a normal ability to elicit the escape behaviour in response to noxious stimuli, whereas *Dmpiezo* is specifically required for the mechanical modality of nociception. To evaluate the possible role of *Dmpiezo* in other modes of larval mechanical sensing, we tested the sensitivity of *Dmpiezo* knockout to gentle touch, which is mediated through ciliated neurons^{17,18}. We observed no defect in the sensitivity of *Dmpiezo* knockout larvae to innocuous gentle touch (Fig. 1g).

A mechanical nociception phenotype was previously observed in mutants of *ppk*, a DEG/ENaC channel⁹ and *painless* (*pain*), a TRPA

¹Department of Cell Biology, Dorris Neuroscience Center, The Scripps Research Institute (TSRI), La Jolla, California 92037, USA. ²Genomic Institute of the Novartis Research Foundation (GNF), San Diego, California 92121, USA.

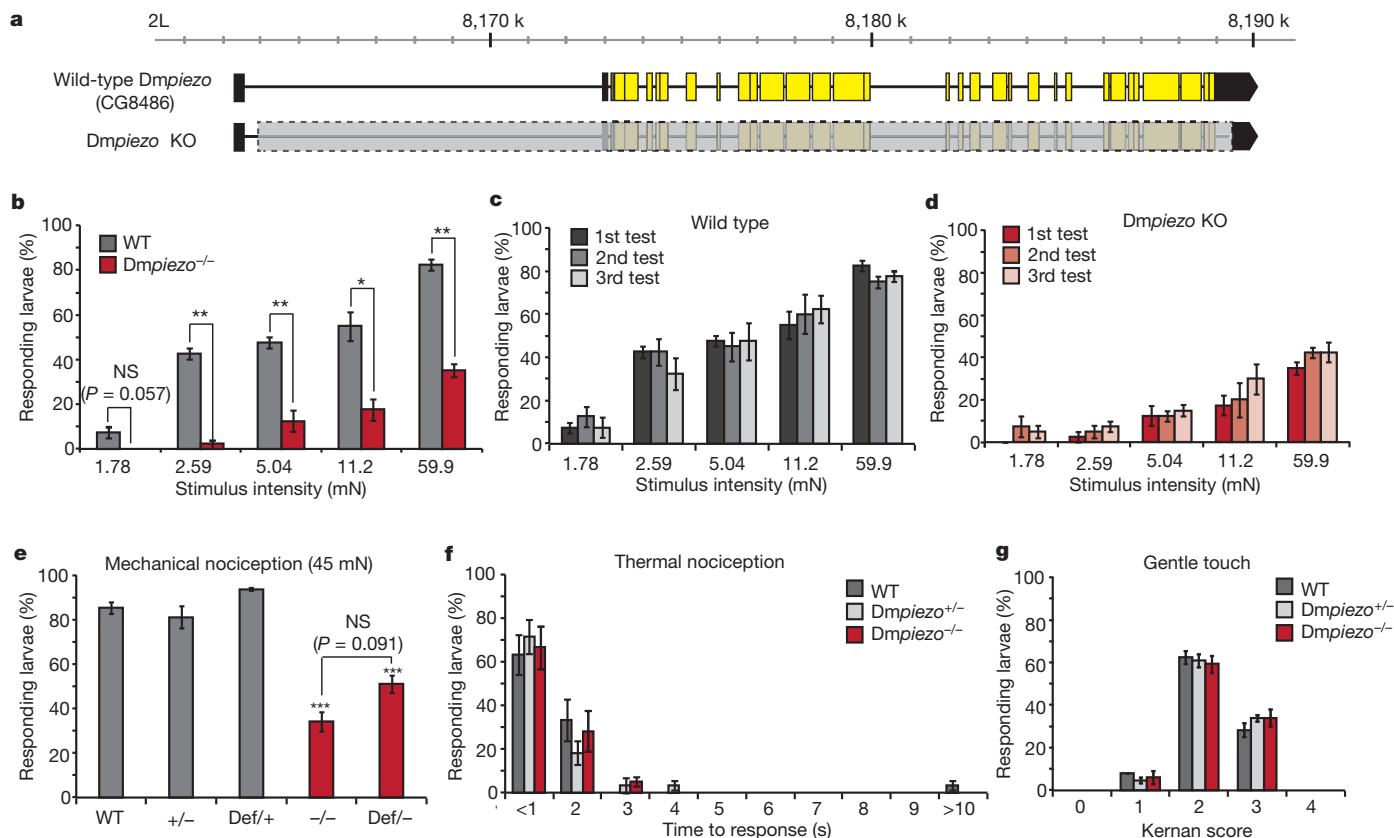


Figure 1 | Mechanical nociception defect in *Dmpiezo* knockout larvae.

a, Genomic map showing wild-type *Dmpiezo* gene (top) and engineered *Dmpiezo* knockout (bottom). Yellow and black boxes represent coding and non-coding exons, respectively. The segment deleted from the left arm of chromosome 2 (2L) in *Dmpiezo* knockout is marked with a grey box. **b**, Mechanical nociception assay using a range of stimulus forces in wild-type (WT) and *Dmpiezo* knockout larvae. $n = 40$ from four independent experiments. * $P < 0.05$, ** $P < 0.01$ from two-tailed paired Student *t*-test.

ion channel¹⁴. The specificity of *Dmpiezo* knockout to mechanical nociception resembles the phenotype of *ppk*, as *pain* is also essential for sensing thermal nociception¹⁴. We therefore tested the role of *Dmpiezo* in *ppk*-positive cells using *ppk*-GAL4, which labels subclasses of multidendritic neurons^{19,20}. The multidendritic neurons are non-ciliated receptor cells that tile the body wall of the larvae and respond

to a variety of external stimuli such as mechanical forces, temperature and light^{9,14,16,21}. We used enhanced (E)GFP driven directly by the regulatory regions of the *ppk* gene (*ppk*-EGFP)²² together with a red fluorescent protein expression in *Dmpiezo*-positive cells to probe *Dmpiezo* and *ppk* co-expression. Indeed, we did observe that all *ppk*-positive cells also expressed *Dmpiezo* (Fig. 2a). Next we used *ppk*-GAL4 to drive the expression of *Dmpiezo* RNA interference (RNAi) to test whether *Dmpiezo* function is specifically required in *ppk*-expressing cells. The restricted knockdown of *Dmpiezo* resulted in a mechanical nociceptive phenotype (Fig. 2b) similar to the phenotype observed in *Dmpiezo* knockout larvae (Fig. 1e). In a complementary approach, we used *ppk*-GAL4-driven expression of *Dmpiezo* complementary DNA in an attempt to rescue the mechanical nociception phenotype of *Dmpiezo* knockout larvae. We used a fusion between DmPiezo and GFP to monitor expression levels in *ppk* cells and DmPiezo localization within the neurons. GFP-DmPiezo fusion protein induces mechanically activated currents in human cell lines, similar to untagged DmPiezo,

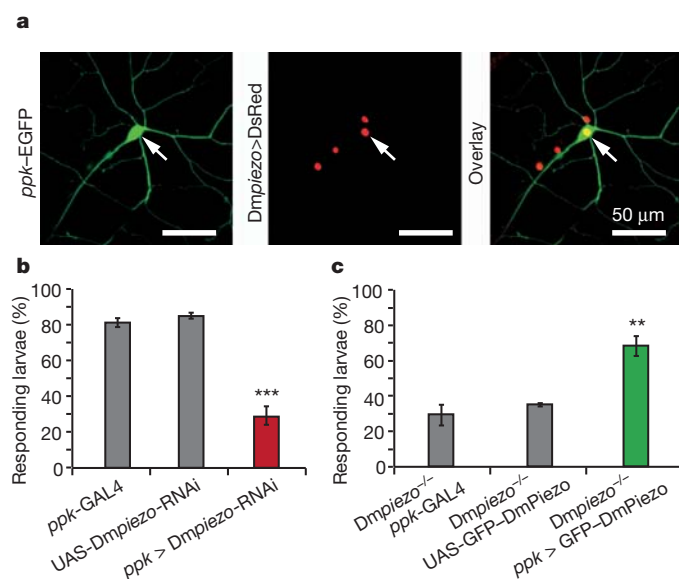


Figure 2 | *Dmpiezo* functions in *ppk*-positive type II sensory neurons.

a, Double fluorescence labelling using *ppk*-EGFP (green) and *Dmpiezo*P-GAL4 that drives the expression of the nucleus targeted UAS-DsRed-NLS (red). A representative high-magnification image shows one *ppk*-positive neuron (arrow). All three *ppk*-positive cells in each hemisegment expressed *Dmpiezo* in all segments. **b**, Mechanical nociception assay with *Dmpiezo* knockdown larvae in *ppk*-expressing cells by *ppk*-GAL4 and UAS-*Dmpiezo*-RNAi. $n > 85$, *** $P < 0.001$. **c**, Mechanical nociception assay in rescued *Dmpiezo* knockout. GFP-DmPiezo was expressed in *ppk*-cells using *ppk*-GAL4 and UAS-GFP-DmPiezo. $n > 60$. ** $P < 0.01$. Error bars indicate mean \pm s.e.m.

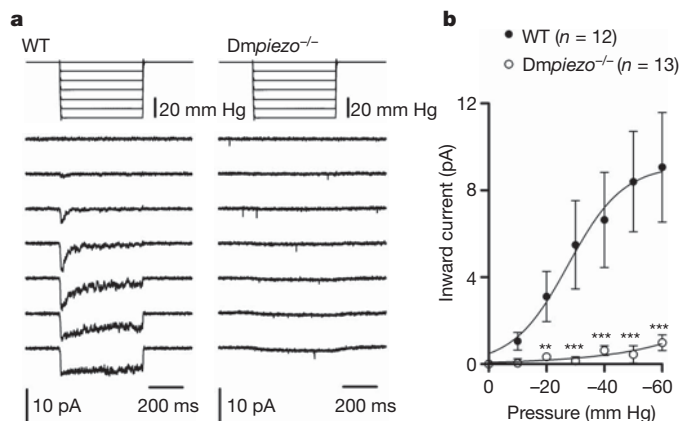


Figure 3 | *Dmpiezo* mediates mechanically activated currents in *ppk*-positive neurons. **a**, Representative currents elicited by negative pipette pressure (0 to -60 mm Hg, $\Delta 10$ mm Hg) in cell-attached configuration at -80 mV in wild type (left) and *Dmpiezo*^{-/-} (right). **b**, Average peak current-pressure relationship of stretch-activated currents in wild type (*n* = 12 cells) and *Dmpiezo*^{-/-} (*n* = 13 cells). Data points are mean \pm s.e.m. fitted with a Boltzmann equation. ***P* < 0.01, ****P* < 0.001, Mann-Whitney test.

confirming functionality (Supplementary Fig. 5a–c). When expressed in *Drosophila*, GFP–DmPiezo fluorescence was present throughout cell bodies, axons and dendritic arborizations of *ppk*-positive neurons (Supplementary Fig. 5d). Importantly, expression of GFP–DmPiezo in *ppk*-positive neurons alone was sufficient to rescue the mechanical nociception defect of *Dmpiezo* knockout larvae (Fig. 2c). These data suggest that *Dmpiezo* functions in *ppk*-positive neurons to mediate mechanical nociception.

To test if the *ppk*-positive neurons respond to mechanical stimuli and if *Dmpiezo* mediates such responses, we performed electrophysiological recordings from isolated cells. Larvae that had GFP labelling in *ppk*-positive neurons were dissociated using enzymatic digestion and mechanical trituration. Plated fluorescent neurons were then tested

using patch-clamp recordings in the cell-attached configuration, and they were stimulated using negative pressure through the recording pipette¹⁰. Stimulating wild-type neurons resulted in a current that was rapidly activated and had a half-maximal activation (P_{50}) of 27.6 ± 7.6 mm Hg (Fig. 3). These currents were not observed in the *Dmpiezo* knockout mutant neurons (Fig. 3). Therefore, *ppk*-positive neurons, which mediate the avoidance response to noxious stimuli, display *Dmpiezo*-dependent, mechanically activated currents.

Silencing of *ppk* cells resulted in complete abolition of noxious mechanosensation (Supplementary Fig. 6), in accord with the severe defect previously observed¹⁶. In contrast, only a moderate deficit is observed upon eliminating or knocking down *ppk* in the same cells⁹, suggesting that there are multiple pathways for mechanical sensing. We tested mechanical nociception in larvae that are deficient in *Dmpiezo* and either *pain* or *ppk* to gain insight into cellular pathways that involve mechanotransduction in these cells. Once again, we used a 45 mN filament, enabling us to monitor both *Dmpiezo*-dependent and independent mechanisms (Fig. 1b). The *Dmpiezo*:*pain* double mutant had a defect that was comparable to each one of the mutants separately, suggesting that *Dmpiezo* and *pain* might function in the same pathway (Fig. 4a). Larvae that are heterozygous for both *Dmpiezo* and *pain* showed a response deficit whereas each one of them separately was normal (Fig. 4b), further demonstrating their role in a common signalling mechanism. Remarkably, combining both *Dmpiezo* and *ppk* knockdowns resulted in a nearly complete abolishment of responses to noxious mechanical stimuli (Fig. 4c). Importantly, responses to noxious temperatures and touch were normal in larvae with both *Dmpiezo* and *ppk* knocked down (Fig. 4d, e). These data indicate that *Dmpiezo* and *ppk* function in two parallel pathways in *ppk*-positive sensory neurons, and that together they constitute the response to noxious mechanical stimuli. There could be many reasons why the mechanically activated currents we observe are entirely dependent on DmPiezo (Fig. 3). This could either be because PPK responds to a different modality of mechanical stimulus or due to the specific experimental settings (for example, level of applied forces, solutions, applied voltage). Future experiments should resolve this issue.

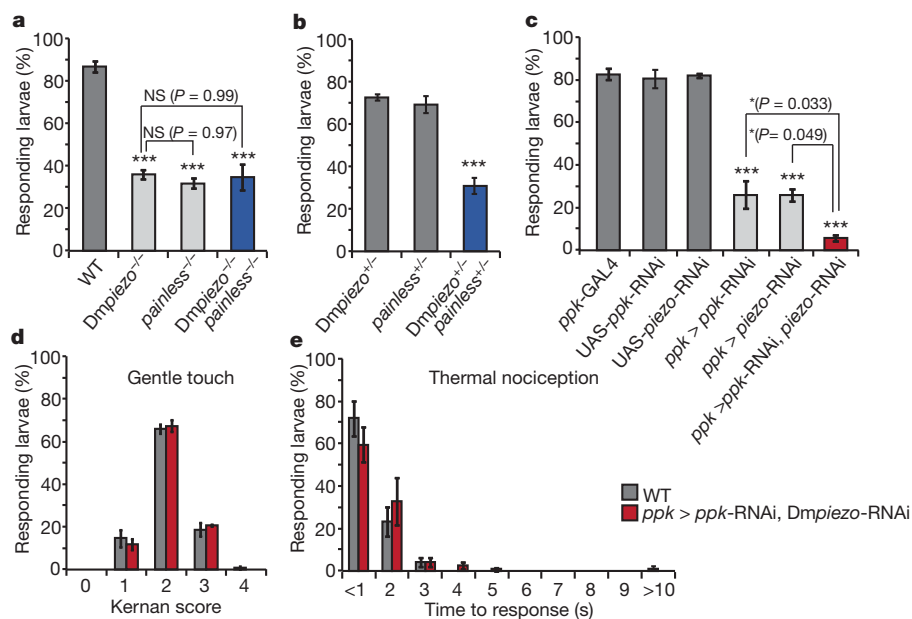


Figure 4 | *Dmpiezo* and *ppk* function in parallel pathways. **a**, Mechanical nociception assay using a 45 mN von Frey filament with double-null mutant of *Dmpiezo* and *painless*. Single-knockout strains were used as controls and the wild-type strain is *w*¹¹¹⁸. *n* > 60. **b**, Mechanical nociception assay on heterozygous larvae for *Dmpiezo* and/or *pain*. *n* (heterozygote *Dmpiezo* knockout) = 74 from three trials, *n* (heterozygote *painless*) = 169 from five trials, *n* (trans-heterozygote) = 166 from five trials. **c**, Mechanical nociception

assay with *ppk* and *Dmpiezo* knockdown. *ppk* and/or *Dmpiezo* RNAi were driven by *ppk*-GAL4. *n* > 90. **P* < 0.05, ****P* < 0.001. **d**, Gentle touch sensitivity assay with *ppk* and *Dmpiezo* knockdown. For details about the Kernan score, see Methods. Wild type is *w*¹¹¹⁸. *n* > 90. **e**, Thermal nociception assay using 45 °C probe with *ppk* and *Dmpiezo* knockdown. *n* > 75. Error bars indicate mean \pm s.e.m.

Using the *Drosophila* model system we have demonstrated that *piezo* is essential for sensing noxious mechanical stimulus *in vivo*. This is the first demonstration that a Piezo family member is essential for mechanotransduction in the whole animal. Indeed, *Dmpiezo* is, to our knowledge, the first eukaryotic excitatory channel component shown to be activated by mechanical force in a heterologous expression system and required for sensory mechanotransduction *in vivo*. *Piezo2* is expressed in mouse dorsal root ganglion neurons that are involved in sensing nociception, and is required for rapidly adapting mechanically activated currents in such isolated neurons¹⁰. This study raises the possibility that mammalian *Piezo2* is also required for mechanical pain transduction *in vivo*. Furthermore, *Drosophila* genetics can now be used to map cellular pathways involved in *piezo*-dependent mechanotransduction in sensory neurons and beyond.

METHODS SUMMARY

Fly stocks. PiggyBacs (PBac{WH}CG8486-f02291, PBac{RB}CG8486-e00109; Exelixis Collection at the Harvard Medical School), *ppk*-GAL4 (Bloomington *Drosophila* Stock Center (BDSC), 32078, 32079), Deficiency (Df(2L)Exel7034/CyO; BDSC, 7807), UAS-*Dmpiezo*-RNAi (National Institute of Genetics, Japan, 8486R-3), UAS-*ppk*-RNAi (Vienna *Drosophila* RNAi Center, 108683), *ppk*-EGFP5 (ref. 22; Y. N. Jan), *painless*¹ (BDSC, 27895).

Generating *Dmpiezo* knockout flies. The *Dmpiezo* knockout fly was generated by FLP-FRT recombination with two PiggyBac lines as described previously¹⁵. The recombined knockout fly was confirmed by PCR (Supplementary Fig. 3). The genetic background was cleaned using meiotic recombination with *w¹¹¹⁸*.

Imaging. Fluorescence in adult fly or larva was detected by Nikon C2 Confocal Laser Point Scanning Microscope, Olympus FluoView500 Confocal Microscope or Olympus AX70 microscope.

Cloning. To clone the enhancer/promoter of the *Dmpiezo* gene, the genomic region between 1.0 kb upstream of the beginning of transcription and the start codon of *Dmpiezo* was amplified by PCR and cloned into the pPTGAL vector. The GFP-*Dmpiezo* construct has three alanines as a linker between the carboxy-terminal GFP and amino-terminal *Dmpiezo*. The construct was cloned in modified pUAST vector to generate transgenic flies and in modified pIRES2-EGFP vector for electrophysiology recordings.

Behavioural assays and statistics. The mechanical nociception was tested as described previously^{9,14,16} using calibrated von Frey filaments. The thermal nociception was tested as described previously¹⁴ using a 45 °C heated metal probe. All error bars represent mean ± s.e.m.

Isolation of *ppk*-positive neurons. Third instar larvae that had GFP labelling in *ppk*-positive neurons were dissected, digested with collagenase and mechanically triturated. The cells were collected by centrifugation and plated on a poly-D-lysine-coated glass coverslip. The fluorescent *ppk*-positive cells were recorded after incubating for 2 h at room temperature (23–25 °C).

Electrophysiology. HEK cells were studied in the whole cell configuration using a polished glass probe for stimulation¹⁰ and *ppk*-positive neurons were stimulated using negative pressure in the cell attached configuration¹⁰.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 July; accepted 20 December 2011.

Published online 19 February 2012.

- Chalfie, M. Neurosensory mechanotransduction. *Nature Rev. Mol. Cell Biol.* **10**, 44–52 (2009).

- Tsunozaki, M. & Bautista, D. M. Mammalian somatosensory mechanotransduction. *Curr. Opin. Neurobiol.* **19**, 362–369 (2009).
- Gillespie, P. G. & Muller, U. Mechanotransduction by hair cells: models, molecules, and mechanisms. *Cell* **139**, 33–44 (2009).
- Delmas, P., Hao, J. & Rodat-Despoix, L. Molecular mechanisms of mechanotransduction in mammalian sensory neurons. *Nature Rev. Neurosci.* **12**, 139–153 (2011).
- Walker, R. G., Willingham, A. T. & Zuker, C. S. A *Drosophila* mechanosensory transduction channel. *Science* **287**, 2229–2234 (2000).
- Li, W., Feng, Z., Sternberg, P. W. & Xu, X. Z. S. A *C. elegans* stretch receptor neuron revealed by a mechanosensitive TRP channel homologue. *Nature* **440**, 684–687 (2006).
- Kang, L., Gao, J., Schafer, W. R., Xie, Z. & Xu, X. Z. S. *C. elegans* TRP family protein TRP-4 is a pore-forming subunit of a native mechanotransduction channel. *Neuron* **67**, 381–391 (2010).
- O'Hagan, R., Chalfie, M. & Goodman, M. B. The MEC-4 DEG/ENaC channel of *Caenorhabditis elegans* touch receptor neurons transduces mechanical signals. *Nature Neurosci.* **8**, 43–50 (2005).
- Zhong, L., Hwang, R. Y. & Tracey, W. D. Pickpocket is a DEG/ENaC protein required for mechanical nociception in *Drosophila* larvae. *Curr. Biol.* **20**, 429–434 (2010).
- Coste, B. *et al.* Piezo1 and Piezo2 are essential components of distinct mechanically activated cation channels. *Science* **330**, 55–60 (2010).
- Coste, B. *et al.* Piezo proteins are pore-forming subunits of mechanically activated channels. *Nature* **483**, <http://dx.doi.org/10.1038/nature10812> (this issue).
- Kim, J. *et al.* A TRPV family ion channel required for hearing in *Drosophila*. *Nature* **424**, 81–84 (2003).
- Gong, Z. *et al.* Two interdependent TRPV channel subunits, Inactive and Nanchung, mediate hearing in *Drosophila*. *J. Neurosci.* **24**, 9059–9066 (2004).
- Tracey, W. D. Jr, Wilson, R. I., Laurent, G. & Benzer, S. *painless*, a *Drosophila* gene essential for nociception. *Cell* **113**, 261–273 (2003).
- Parks, A. L. *et al.* Systematic generation of high-resolution deletion coverage of the *Drosophila melanogaster* genome. *Nature Genet.* **36**, 288–292 (2004).
- Hwang, R. Y. *et al.* Nociceptive neurons protect *Drosophila* larvae from parasitoid wasps. *Curr. Biol.* **17**, 2105–2116 (2007).
- Kernan, M., Cowan, D. & Zuker, C. Genetic dissection of mechanosensory transduction: mechanoreception-defective mutations of *Drosophila*. *Neuron* **12**, 1195–1206 (1994).
- Caldwell, J. C., Miller, M. M., Wing, S., Soll, D. R. & Eberl, D. F. Dynamic analysis of larval locomotion in *Drosophila* chordotonal organ mutants. *Proc. Natl Acad. Sci. USA* **100**, 16053–16058 (2003).
- Adams, C. M. *et al.* Ripped Pocket and Pickpocket, novel *Drosophila* DEG/ENaC subunits expressed in early development and in mechanosensory neurons. *J. Cell Biol.* **140**, 143–152 (1998).
- Ainsley, J. A. *et al.* Enhanced locomotion caused by loss of the *Drosophila* DEG/ENaC protein Pickpocket1. *Curr. Biol.* **13**, 1557–1563 (2003).
- Xiang, Y. *et al.* Light-avoidance-mediating photoreceptors tile the *Drosophila* larval body wall. *Nature* **468**, 921–926 (2010).
- Grueber, W. B., Ye, B., Moore, A. W., Jan, L. Y. & Jan, Y. N. Dendrites of distinct classes of *Drosophila* sensory neurons show different capacities for homotypic repulsion. *Curr. Biol.* **13**, 618–626 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. N. Jan of the University of California San Francisco for providing *ppk*-EGFP5. Research was support by the National Institutes of Health and Novartis Research Foundation. S.E.K. and A.C. are supported by the Skaggs Institute.

Author Contributions S.E.K. conducted most experiments. B. Coste performed the electrophysiology experiments shown in Fig. 3 and Supplementary Fig. 5. A.C. performed the fly electrophysiology experiments shown in Supplementary Fig. 4. S.E.K., A.P. and B. Cook designed experiments and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.C. (bcook@scripps.edu) or A.P. (apatapou@gnf.org).

METHODS

Fly stocks. We used the following stocks: PiggyBacs (PBac{WH}CG8486-f02291, PBac{RB}CG8486-e00109, Exelixis Collection at the Harvard Medical School), *ppk*-GAL4 (Bloomington *Drosophila* Stock Center (BDSC), 32078, 32079), Deficiency (Df(2L)Exel7034/CyO, BDSC, 7807), UAS-*Dmpiezo*-RNAi (National Institute of Genetics, Japan, 8486R-3), UAS-*ppk*-RNAi (Vienna *Drosophila* RNAi Center, 108683), *ppk*-EGFP5 (ref. 22; Y. N. Jan), *painless*¹ (BDSC, 27895) and UAS-DsRed-NLS (J. W. Posakony). The following stocks were from BDSC: UAS-GFP, UAS-CD8::GFP, CyO-GFP, *w*¹¹¹⁸ and *Canton-S*.

Engineering *Dmpiezo* knockout flies. The *Dmpiezo* knockout fly was generated as described in previously described¹⁵. Two PiggyBac lines that carry the FRT sequence were selected for FLP-FRT recombination. PBac{WH}CG8486-f02291 is inserted in the first intron and PBac{RB}CG8486-e00109 in the 3' untranslated region (UTR) of the *Dmpiezo* gene. After FLP-FRT recombination, 20 kb of the *Dmpiezo* gene, including all 31 coding exons, was removed and replaced with 7 kb of PiggyBac insertion that contained the FRT sequence and white gene. The recombined knockout fly was confirmed by PCR reactions (Supplementary Fig. 2). The genetic background was cleaned using meiotic recombination with *w*¹¹¹⁸.

Molecular biology. To clone the enhancer/promoter of the *Dmpiezo* gene, the genomic region between 1.0 kb upstream of the beginning of transcription and the start of the *Dmpiezo* coding region was amplified by PCR using forward primer, 5'-ATCTGGCGGCCGCTATCTATTTTAACTAGTGGAAGTCT-3' and reverse primer, 5'-TTACTGGTACCATGGATGCCTCCGCGCGCGTTC TCCTCCAG-3'. The amplified sequence was cloned into pPTGAL vector (*Drosophila* Genomic Resource Center, 1225) using NotI and KpnI sites and the sequence was verified.

For rescue experiments, *Dmpiezo* cDNA was amplified from the plasmid reported in ref. 11, using forward primer 5'-TATTAGCGGCCGAGTCTTCA GCTATGCGTGCATGGTG-3' and reverse primer 5'-TAATTCGGTCCGTTAT TGCGGTTGCTGTGGTGCAGTTGCTCCGG-3' and cloned into a modified pUAST vector using NotI and RsrII. NotI restriction enzyme site was used as a linker by providing three alanine residues between EGFP and DmPiezo. The order of sequences in the pUAST vector is the following: UAS-kozak-EGFP-3×(Ala)-DmPiezo. To generate transgenic flies, DNA was injected into the isogenized *w*¹¹¹⁸ embryos along with transposase Δ2-3. For the electrophysiology experiment, EGFP-DmPiezo was cloned into mammalian expression vector with CMV promoter.

Behaviour assays. Mechanical nociception was tested as described previously^{9,14,16} using calibrated von Frey filaments. Thermal nociception was tested as described previously¹⁴ using a calibrated heated metal probe. For both nociception assays, the number of larvae that showed at least one 360° rotation was counted for each trial. The gentle touch assay was performed and each stimulated larva was scored as described previously¹⁷. 0 = no response, 1 = hesitates, 2 = turns or withdraws anterior segments, 3 = single reverse contractile wave, and 4 = multiple waves. For all behaviour assays each third instar larva was stimulated only once. All data were generated from at least three trials.

The von Frey filaments for larvae behaviour experiments were modified from Touch-Test sensory Evaluator (North Coast Medical) or from monofilament fishing lines. Each monofilament was cut to a length of 18 mm, glued into a pipette tip so that 9 mm of it protruded and mounted on a hand manipulator with a 90° angle. Each von Frey filament was calibrated as described previously⁹. The force of each von Frey stimulator was determined by measuring the weight upon filament bending and converting the value into the force: force (mN) = mass (g) × gravity acceleration constant (g; 9.8). Each stimulator was calibrated 15 times and its mean

value was used in figures. The calibrated forces (mean ± s.e.m.) of each stimulator are as follows (in mN): 1.78 ± 0.15, 2.59 ± 0.15, 5.04 ± 0.19, 11.2 ± 0.66 and 59.9 ± 1.79.

Fluorescence imaging. For identifying tissues or cells expressing fluorescence by the *Dmpiezo* promoter, both adult flies and third instar larvae carrying *Dmpiezo*P-GAL4 and UAS-GFP, or UAS-CD8::GFP, were dissected or whole-mounted. For double fluorescent labelling in multidendritic neurons, second instar larvae carrying *ppk*-EGFP, *Dmpiezo*P-GAL4 and UAS-DsRed were whole-mounted. For imaging *ppk*-cells expressing GFP-DmPiezo, third instar larvae carrying *ppk*-GAL4 and UAS-GFP-DmPiezo were whole-mounted. Fluorescence images were obtained either by Nikon C2 Confocal Laser Point Scanning Microscope, Olympus FluoView500 Confocal Microscope or Olympus AX70 microscope.

Isolation of larvae *ppk*-positive neurons. In both wild-type and *Dmpiezo* knockout larvae, *ppk*-positive neurons were fluorescently labelled by *ppk*-EGFP, which is a direct fusion of *ppk* genomic regulatory regions with EGFP. Third instar larvae were dissected in M3 media containing 10% heat inactivated FBS. Each larva was cut twice and its internal organs were removed. The cleaned body wall was treated with 5 mg ml⁻¹ collagenase type IV at 25 °C for 1 h in serum-free M3 media and washed with serum containing M3 media. The enzyme-treated body wall was triturated with fire-polished Pasteur pipettes in M3 media with 2 mM EGTA and 10% FBS. The cuticle and debris were removed by centrifugation at 40g and the small size cells including neurons were collected by centrifugation at 360g for 10 min. The cell pellet was resuspended with serum containing M3 media and plated into a poly-D-lysine-coated coverslip in a small droplet. After 2 h of incubation at room temperature (23–25 °C), the coverslips were transferred to the electrophysiology rig for recording.

Electrophysiology. For whole-cell recordings in HEK293T cells, patch pipettes had resistances of 2–3 MΩ when filled with an internal solution consisting of (in mM) 133 CsCl, 10 HEPES, 5 EGTA, 1 CaCl₂, 1 MgCl₂, 4 MgATP and 0.4 Na₂GTP (pH adjusted to 7.3 with CsOH). The extracellular solution consisted of (in mM) 130 NaCl, 3 KCl, 1 MgCl₂, 10 HEPES, 2.5 CaCl₂, 10 glucose (pH adjusted to 7.3 with NaOH). Mechanical stimulation was achieved using a fire-polished glass pipette (tip diameter 3–4 μm). The probe had a velocity of 1 μm ms⁻¹ during the ramp segment of the command for forward motion and the stimulus was applied for 150 ms.

For cell-attached recordings in *ppk*-positive dissociated neurons, patch pipettes had resistances of 3–3.5 MΩ when filled with a solution consisting of (in mM) 130 NaCl, 5 KCl, 10 HEPES, 1 CaCl₂, 1 MgCl₂, 10 TEA-Cl (pH 7.3 with NaOH). External solution used to zero the membrane potential consisted of (in mM) 140 KCl, 10 HEPES, 1 MgCl₂, 10 glucose (pH 7.3 with KOH). Membrane patches were stimulated with brief negative pressure pulses through the recording electrode using a Clampex controlled pressure clamp HSPC-1 device (ALA-scientific). Stretch-activated channels were recorded at a holding potential of -80 mV with pressure steps from 0 to -60 mm Hg (-10 mm Hg increments). Current-pressure relationships were fitted with a Boltzmann equation of the form: $I(P) = (1 + \exp(-(P - P_{50})/s))^{-1}$, where I is the peak of stretch-activated current at a given pressure, P is the applied patch pressure (in mm Hg), P_{50} is the pressure value that evoked a current value which is 50% of I_{\max} , and s reflects the current sensitivity to pressure.

All experiments were performed at room temperature. Currents were sampled at 50 or 20 kHz and filtered at 5 or 2 kHz. Voltages were not corrected for a liquid junction potential. Leak currents before mechanical stimulations were subtracted off-line from the current traces.

Piezo proteins are pore-forming subunits of mechanically activated channels

Bertrand Coste^{1*}, Bailong Xiao^{1*}, Jose S. Santos², Ruhma Syeda², Jörg Grandl^{1†}, Kathryn S. Spencer¹, Sung Eun Kim¹, Manuela Schmidt¹, Jayanti Mathur³, Adrienne E. Dubin¹, Mauricio Montal² & Ardem Patapoutian^{1,3}

Mechanotransduction has an important role in physiology. Biological processes including sensing touch and sound waves require as-yet-unidentified cation channels that detect pressure. Mouse Piezo1 (MmPiezo1) and MmPiezo2 (also called Fam38a and Fam38b, respectively) induce mechanically activated cationic currents in cells; however, it is unknown whether Piezo proteins are pore-forming ion channels or modulate ion channels. Here we show that *Drosophila melanogaster* Piezo (DmPiezo, also called CG8486) also induces mechanically activated currents in cells, but through channels with remarkably distinct pore properties including sensitivity to the pore blocker ruthenium red and single channel conductances. MmPiezo1 assembles as a ~1.2-million-dalton homo-oligomer, with no evidence of other proteins in this complex. Purified MmPiezo1 reconstituted into asymmetric lipid bilayers and liposomes forms ruthenium-red-sensitive ion channels. These data demonstrate that Piezo proteins are an evolutionarily conserved ion channel family involved in mechanotransduction.

Mechanically activated currents have been described in various mammalian cells, including inner ear hair cells¹, somatosensory dorsal root ganglion neurons², vascular smooth muscle cells³ and kidney primary epithelia⁴. Most of these mechanically activated currents are cationic with Ca²⁺ permeability, leading to a search for cation channels able to convert mechanical forces into such currents. Few mechanically activated channels have been described so far^{5–7}; however, none of the candidates has been shown convincingly to mediate the physiologically relevant non-selective cationic mechanically activated currents in mammals.

MmPiezo1 was recently identified as a protein required for mechanically activated currents in a mammalian cell line. Expressing MmPiezo1 or related MmPiezo2 in a variety of mammalian cell lines induces mechanically activated cationic currents⁸. MmPiezo1-induced currents are inhibited by GsMTx4 (*Grammostola spatulata* mechanotoxin 4), a peptide widely used to study mechanically activated channels⁹. MmPiezo1 and MmPiezo2 contain over 30 putative transmembrane domains and do not resemble known ion channels or other protein classes. Piezo proteins could be non-conducting subunits of cationic ion channels required for proper expression or for modulating channel properties^{6,10,11}. Alternatively, Piezo proteins may define a novel class of ion channels involved in mechanotransduction.

Mechanosensitivity of DmPiezo

Piezo sequences are present in the genomes of many animal, plant and other eukaryotic species. Functional analysis of Piezo proteins from phylogenetically distant species could demonstrate a conserved role of these proteins in mechanotransduction; furthermore, a comparative analysis of mechanically activated currents could elucidate unique pore properties of channels induced by Piezo proteins from distinct species. We focused on the apparently single member of *D. melanogaster* Piezo (DmPiezo), as this invertebrate species is widely used to study mechanotransduction using genetic approaches^{12–16}. DmPiezo is 24% identical to mammalian Piezo proteins, with sequence conservation

throughout the length of the proteins (Supplementary Fig. 1). We cloned the full-length DmPiezo complementary DNA into pIRES2-EGFP vector. We recorded mechanically activated currents from fluorescent HEK293T cells expressing DmPiezo-pIRES2-EGFP by applying force to the cell surface while monitoring transmembrane currents at constant voltage using patch-clamp recordings in the whole-cell configuration^{2,17,18}. DmPiezo, but not mock-transfected cells, showed large mechanically activated currents (Fig. 1a, b). These currents have a time constant of inactivation τ of 6.2 ± 0.3 ms ($n = 32$ cells) at -80 mV when fitted with mono-exponential function, which is faster than observed for MmPiezo1 (~16 ms) and more comparable to MmPiezo2 (~7 ms)⁸. Similar to its mammalian counterparts, DmPiezo mechanically activated currents are characterized by a linear current-voltage (I - V) relationship with a reversal potential around 0 mV, consistent with it mediating a non-selective cationic conductance (Fig. 1c). We further characterized DmPiezo-induced currents in HEK293T cells in response to negative pressure pulses applied through the recording pipette in the cell-attached mode, an alternative mechanosensitivity assay. Overexpression of DmPiezo induced stretch-activated currents (Fig. 1d, e) with a pressure for half-maximal activation (P_{50}) of -31.8 ± 2.8 mm Hg (Fig. 1f), similar to the P_{50} calculated for MmPiezo1-induced currents (~30 mm Hg)⁸. Therefore, mechanosensitivity of the Piezo family is conserved in invertebrates. We demonstrate the physiological relevance of DmPiezo *in vivo* in an accompanying paper¹⁹.

Pore properties of Piezo proteins

We next compared fundamental permeation properties of MmPiezo1 and DmPiezo. Ruthenium red, a polycationic pore blocker of TRP channels^{20,21}, blocks MmPiezo1- and MmPiezo2-induced mechanically activated currents⁸. We found that ruthenium red is a voltage-dependent blocker of MmPiezo1, with an IC₅₀ value of 5.4 ± 0.9 μ M at -80 mV (Fig. 2a–c): at a concentration of 30 μ M, extracellular ruthenium red inhibited inward mechanically activated currents without affecting

¹Department of Cell Biology, Dorris Neuroscience Center, The Scripps Research Institute, La Jolla, California 92037, USA. ²Section of Neurobiology, Division of Biological Sciences, University of California San Diego, La Jolla, California 92093, USA. ³Genomic Institute of the Novartis Research Foundation, San Diego, California 92121, USA. [†]Present address: Department of Neurobiology, Duke University Medical Center, Durham, North Carolina 27710, USA.

*These authors contributed equally to this work.

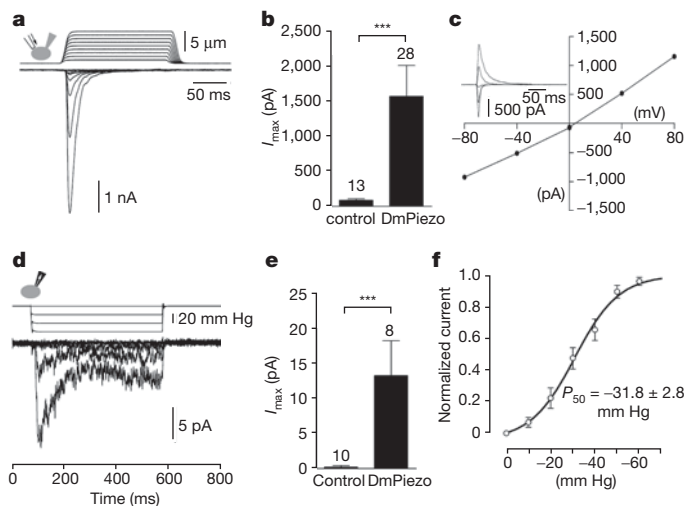


Figure 1 | Human cells expressing *Drosophila* Piezo (DmPiezo) show large mechanically activated currents. **a–f**, Mechanically activated currents of DmPiezo-expressing HE293T cells recorded in the whole-cell (**a–c**) or cell-attached (**d–f**) configuration. **a**, Representative traces of mechanically activated inward currents at -80 mV in DmPiezo-transfected cells subjected to a series of mechanical steps in $1\ \mu\text{m}$ increments. **b**, Average maximal current amplitude of mechanically activated inward currents at -80 mV. **c**, Representative I – V relationship of mechanically activated currents in DmPiezo-transfected cells. The inset shows mechanically activated currents evoked at holding potentials ranging from -80 to $+80$ mV. **d**, Representative currents elicited by negative pipette pressure (0 to -60 mm Hg, $\Delta 20$ mm Hg) in DmPiezo-transfected cells. **e**, Average maximal current amplitude of stretch-activated currents at -80 mV. **f**, I_{max} normalized current–pressure relationship of stretch-activated currents recorded at -80 mV in DmPiezo-transfected cells ($n = 8$ cells) and fitted with a Boltzmann equation. P_{50} is the average of P_{50} values determined for individual cells. Bars represent mean \pm s.e.m. and the number of cells tested is shown above bars. *** $P < 0.001$, Mann–Whitney U -test.

outwards currents. Such voltage dependence is a characteristic of open channel block. A high concentration of ruthenium red ($50\ \mu\text{M}$) included in the pipette solution in the whole-cell configuration showed no evidence of block, as large mechanically activated currents still displayed a linear I – V relationship (Supplementary Fig. 2). These results suggest that ruthenium red blocks the pore of MmPiezo1-induced mechanically activated channels from the extracellular side. Notably, DmPiezo-induced mechanically activated currents were insensitive to ruthenium red concentrations that potently blocked MmPiezo1-induced currents (Fig. 2d,e). Together, these results demonstrate that overexpression of DmPiezo or MmPiezo1 gives rise to mechanically activated channels with distinct channel properties.

Next, we set out to determine the single channel conductance (γ) of mechanically activated channels induced by Piezo proteins by using negative-pressure stimulations of membrane patches in cell-attached mode. Figure 3 shows the single mechanically activated channel properties of MmPiezo1 or DmPiezo. Openings of stretch-activated channels showed a marked difference in amplitude of single channel currents (Fig. 3a), as determined from the single channel I – V relationship for MmPiezo1 and DmPiezo (Fig. 3b,c). Linear regression of these I – V relationships resulted in slope-conductance values in these recording conditions of 29.9 ± 1.9 and 3.3 ± 0.3 pS for MmPiezo1- and DmPiezo-induced mechanically activated currents, respectively ($n = 7$ and 5 cells; mean \pm s.e.m.). Therefore, DmPiezo-dependent channels are ninefold less conductive than MmPiezo1-dependent channels.

MmPiezo1 oligomerization

The pore of most ion channels is formed by the assembly of transmembrane domains from distinct subunits (for example, voltage-gated K^+ channels, ligand-gated ion channels) or structurally repetitive domains within a large protein (for example, voltage-gated Na^+ and

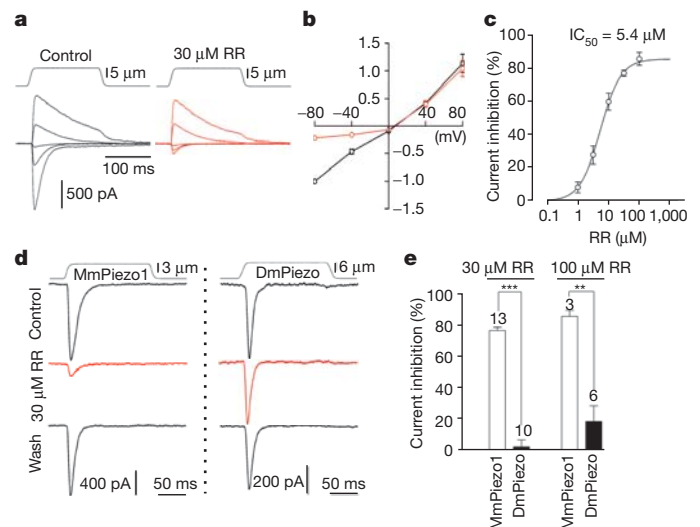


Figure 2 | Ruthenium red is a channel pore blocker of MmPiezo1- but not DmPiezo-induced currents. **a**, Representative traces of mechanically activated currents in MmPiezo1-transfected cells evoked at holding potentials ranging from -80 to $+80$ mV before (left panel) and during perfusion of $30\ \mu\text{M}$ of ruthenium red (right panel, red traces). **b**, Average I – V relationship of mechanically activated currents in MmPiezo1-transfected cells ($n = 7$ cells) before (black symbols) and during (red symbols) perfusion of $30\ \mu\text{M}$ ruthenium red. Currents were normalized to the value of control current evoked at -80 mV for each individual cell. **c**, Concentration–inhibition curve for ruthenium red (RR) on mechanically activated currents evoked at -80 mV in MmPiezo1-transfected cells and fitted with a Boltzmann equation. Each data point is the mean \pm s.e.m. of 3–13 observations. **d**, Representative traces of Piezo-dependent mechanically activated currents evoked at -80 mV in the presence of ruthenium red. **e**, Blocking effect of ruthenium red on Piezo-dependent mechanically activated currents evoked at -80 mV. Bars represent mean \pm s.e.m. and the number of cells tested is shown above the bars. ** $P < 0.01$; *** $P < 0.001$; unpaired t -test.

Ca^{2+} channels). As Piezo proteins lack repetitive transmembrane motifs presumably they oligomerize to form ion channels. To test this hypothesis, we determined the number of subunits in Piezo complexes by expressing GFP–MmPiezo1 fusion proteins in *Xenopus laevis* oocytes, imaging individual spots with total internal reflection microscopy (TIRF), and counting discrete photobleaching steps (Fig. 4a, b and ref. 22). Amino-terminal GFP–MmPiezo1 functionality was confirmed by overexpression in HEK293T cells (Supplementary Fig. 3). We used several GFP fusion constructs of ion channels with known stoichiometry as controls: voltage-gated Ca^{2+} channel ($\alpha 1\text{E}$ -GFP; monomer), NMDA (N -methyl- D -aspartate) receptor (NR1 co-expressed with NR3A-GFP; dimer of dimers) and cyclic nucleotide-gated (CNG) channel (XfA4-GFP; tetramer)²². We found that complexes of MmPiezo1 frequently exhibited at most four photobleaching steps, consistent with the idea that Piezo proteins homo-multimerize. Fluorescent MmPiezo1 (or CNG) complexes exhibiting bleaching in fewer than four steps can be explained by non-functional GFP or pre-bleached GFP²² or general bias against noisier multi-step traces during data analysis (see Methods). Histograms of the number of photobleaching steps observed for MmPiezo1 complexes were comparable to histograms obtained from tetrameric CNG channels (Fig. 4c). These results suggest that in living cells, Piezo proteins can assemble as homo-multimers.

We further characterized Piezo proteins biochemically by heterologously expressing and purifying MmPiezo1 carboxy-terminally fused with a glutathione S -transferase (MmPiezo1-GST). Functionality of MmPiezo1-GST was confirmed by overexpression in HEK293T cells (Supplementary Fig. 3). We observed a protein band at a position near the 260-kDa protein marker on a Coomassie-blue-stained denaturing protein gel (Supplementary Fig. 4a). Western blot with a GST (*Schistosoma japonicum* form) antibody (Supplementary Fig. 4b) or

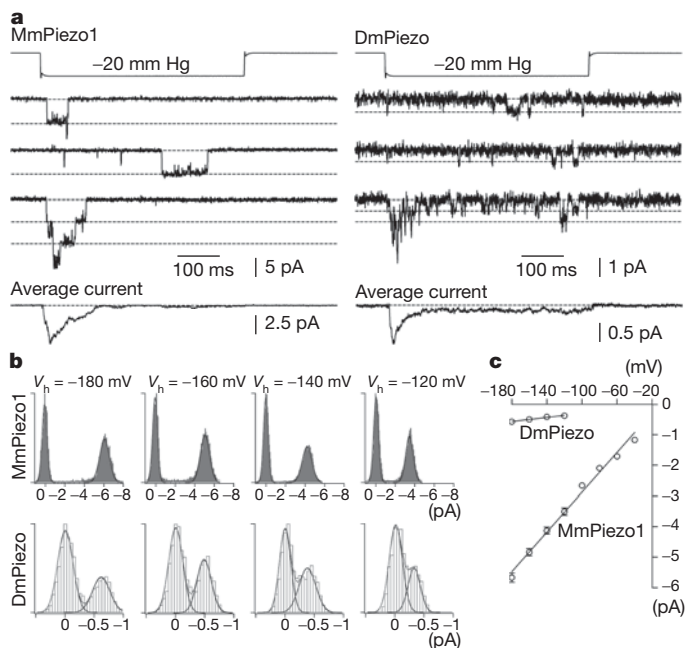


Figure 3 | MmPiezo1- and DmPiezo-induced stretch-activated channels have different conductances. **a**, Representative Piezo-dependent stretch-activated channel openings elicited at -180 mV. Bottom traces represent average of 40 individual recording traces. **b**, All-point histograms of single channel opening events (average of 10 and 20 individual events for MmPiezo1 and DmPiezo, respectively) at different holding potentials (V_h). **c**, Average I - V relationships of stretch activated single channels in MmPiezo1 and DmPiezo transfected cells ($n = 7$ and 5 cells, respectively; mean \pm s.e.m.). Single channel amplitude was determined as the amplitude difference in Gaussian fits as shown in **b**.

a MmPiezo1-specific antibody⁸ (Fig. 4) confirmed the presence of MmPiezo1-GST in the MmPiezo1-GST sample. Using native gel electrophoresis and Coomassie blue staining, we detected a prominent protein band at a position near the 1,236 kDa protein marker only in the MmPiezo1-GST sample (Fig. 4d). Western blot using MmPiezo1 antibody confirmed that this major band contains MmPiezo1 (Fig. 4e). These data indicate that the purified MmPiezo1-GST protein complex has a molecular weight of about 1.2 million Da, four times the predicted molecular weight of a single MmPiezo1-GST polypeptide (318 kDa). Next, we asked whether any endogenous proteins are present in this MmPiezo1-containing complex. Mass spectrometry of the ~ 1.2 million Da protein complex mainly detected peptides derived from MmPiezo1-GST, but not from other endogenous membrane proteins. Although several non-transmembrane proteins were also detected, most of them were also present in the control sample, indicating an absence of specific interacting proteins in the complex (Supplementary Table 1). Moreover, mass spectrometry of the whole purified solution samples before gel electrophoresis confirmed that no other ion channel protein was detected (Supplementary Table 2). This indicates that MmPiezo1 is not tightly associated with any endogenous pore-forming protein.

To examine further whether this Piezo complex is indeed a tetramer, we treated the purified MmPiezo1-GST protein with the crosslinker formaldehyde and subjected the samples to denaturing gel electrophoresis and western blotting. Formaldehyde-treated samples contained three major additional higher-order Piezo-containing bands, with longer treatments increasing the prominence of the higher bands (Fig. 4f). The distribution of the bands on the 3–8% gradient gel suggests that the four bands correspond to monomer, dimer, trimer and tetramer of MmPiezo1-GST (Fig. 4f). The observation that MmPiezo1 is crosslinked by formaldehyde, a crosslinker with a relative short spacer arm (2.3–2.7 Å), suggests that the subunits form a tetramer.

It is possible that MmPiezo1 oligomers associate with other proteins; however, such an association might not withstand the GST purification step. To probe this, we performed paraformaldehyde (PFA) crosslinking experiments on living cells before the purification procedure. On a native gel, the MmPiezo1-GST complex purified from PFA-treated cells also migrated to the position near the 1,236 kDa protein marker, similar to the sample from untreated cells (Fig. 4g). On a denaturing gel, on-cell PFA treatment resulted in four distinct MmPiezo1-specific bands, similar to results of formaldehyde treatment on the purified complex (Fig. 4h). This suggests that MmPiezo1 is not tightly associated with other proteins large enough to alter discernibly its size on denaturing gels, and confirms the results from mass spectrometry analysis. However, our crosslinking studies with PFA might miss weak interactors with MmPiezo1. Regardless, together with the results obtained from single-molecule photobleaching analysis in living cells, our biochemical data suggest that MmPiezo1 forms a homomultimeric ion channel, most likely as a homotetramer.

MmPiezo1 reconstitution in lipid bilayers

Finally, to assess whether Piezo proteins were sufficient to recapitulate the channel properties recorded from Piezo-overexpressing cells, we reconstituted purified MmPiezo1 proteins into lipid bilayers in two distinct configurations: droplet interface lipid bilayers (DIBs) assembled from two monolayers^{23–25} (Fig. 5a–e and l–q) and proteoliposomes²⁶ (Fig. 5f–h). In the first configuration, MmPiezo1 was reconstituted into asymmetric bilayers that mimic the cellular environment: the extracellular-facing lipid monolayer is predominantly neutral whereas the intracellular-facing leaflet is negatively charged²⁷. In contrast, the lipid composition of the bilayer in the second configuration is uniform.

In the DIBs setting, representative segments from a 6-min recording obtained at -100 mV show brief, discrete channel openings (Fig. 5a, b) blocked by addition of $50 \mu\text{M}$ ruthenium red to the neutral facing compartment (Fig. 5c). In contrast, no effect was observed when ruthenium red was introduced into the negative-facing compartment (not shown). We detected efficient block of channel activity even at $5 \mu\text{M}$ ruthenium red (not shown). The asymmetric accessibility of ruthenium red block of reconstituted channels agrees with the data obtained from MmPiezo1-overexpressing HEK293T cells (Fig. 2 and Supplementary Fig. 2), thereby establishing the fidelity of the assays and validating MmPiezo1 protein as an authentic ion channel. The Piezo currents exhibit ohmic behaviour; records displayed at higher resolution (Fig. 5b) clearly demonstrate the occurrence of unitary events with γ values obtained from conductance histograms of 118 ± 15 pS and 80 ± 6 pS ($n = 6$) in symmetric 0.5 M KCl from the negative and positive branches of I - V plots, respectively (Fig. 5d, e).

A similar pattern of activity was obtained from MmPiezo1 reconstituted in asolectin liposomes²⁶ (Fig. 5f–k). A selection of recordings shows the presence of two channels in the membrane which reside predominantly in the open state (Fig. 5f, g), as discerned in a higher time resolution display (Fig. 5k). These recordings were obtained in the presence of $50 \mu\text{M}$ ruthenium red inside the recording pipette, to ensure functional selection of a single population of MmPiezo1 channels facing the ruthenium-red-free compartment. MmPiezo1 in asolectin proteoliposomes under these conditions (symmetric 0.2 M KCl) exhibits a $\gamma = 110 \pm 10$ pS at $V = -100$ mV and 80 ± 5 pS at $V = 100$ mV (Fig. 5h–j) ($n = 8$). Finally, reconstitution of control samples purified from non-transfected cells as well as heat-denatured purified MmPiezo1-GST into either bilayer systems under otherwise identical conditions failed to reproduce this pattern of channel activity (not shown).

We then tested the ability of the reconstituted MmPiezo1 to conduct sodium (Fig. 5l–q). Initially, single channel currents were recorded from asymmetric bilayers in symmetric 0.2 M KCl; $\gamma = 58 \pm 5$ pS (Fig. 5l, o). Subsequent addition of 0.2 M NaCl in the presence of 0.2 M KCl increased the unitary conductance of reconstituted channels to 95 ± 5 pS (Fig. 5m, p) while retaining sensitivity to ruthenium red

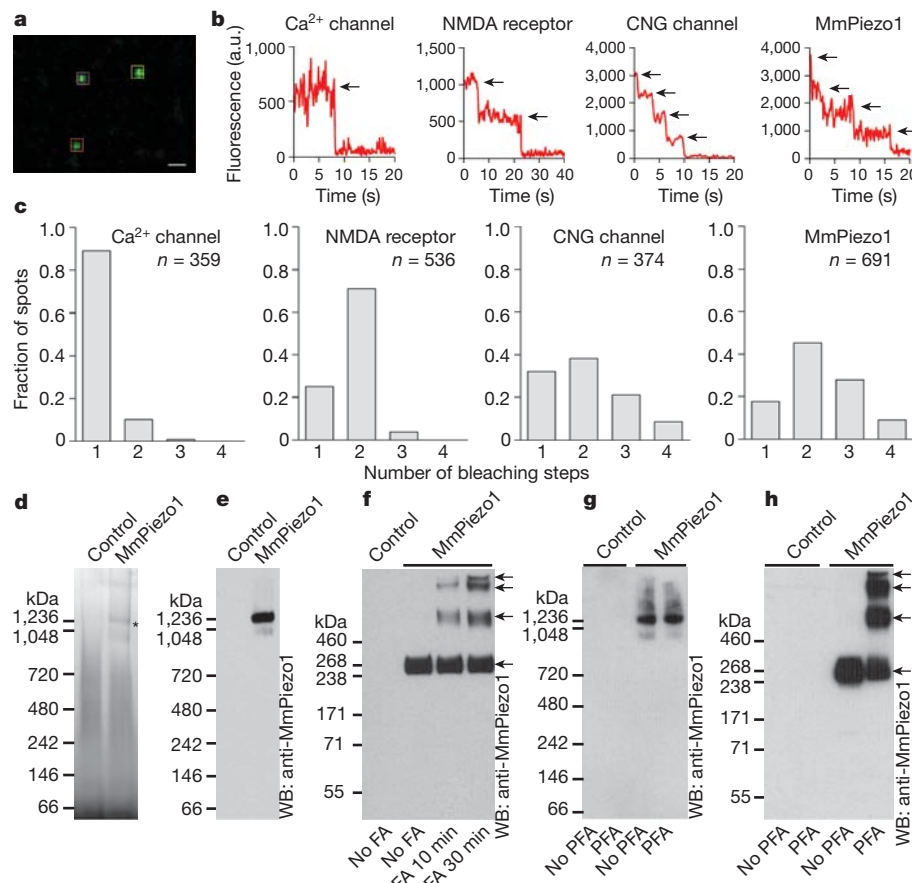


Figure 4 | MmPiezo1 forms homo-oligomers. **a**, Representative image of an acquired sequence showing three selected GFP-MmPiezo1 spots in the cell membrane. Levels were adjusted for clarity. Scale bar, 0.8 μm . **b**, Representative traces of fluorescence intensities of indicated single GFP-fusion constructs. Black arrows indicate photobleaching steps. **c**, Histograms of the average number of bleaching steps observed in ten or more movies from four or more oocytes of single fluorescent complexes of indicated constructs. **d, e**, Indicated samples purified and separated on native gels and visualized by Coomassie staining (**d**) or western blotting (**e**). Asterisk in **d** indicates a protein band

specifically present in the MmPiezo1 sample. **f**, Purified MmPiezo1-GST proteins treated with or without formaldehyde (FA) with the indicated time period, separated on a denaturing gel and detected with the anti-MmPiezo1 antibody. Sample purified from cells without transfection served as a negative control. **g, h**, MmPiezo1-GST-transfected HEK293T cells or untransfected cells treated with or without 0.25% PFA for 10 min. The crosslinked MmPiezo1-GST proteins were purified and separated on native gel (**g**) or denaturing gels (**h**), followed by western blotting. Panels **d-h** are representatives of at least three independent experiments.

block (Fig. 5n, q). These results confirm that these channels conduct both sodium and potassium as would be expected from a cationic non-selective channel. This assertion was further substantiated by recording MmPiezo1 currents from proteoliposomes under bi-ionic conditions (0.2 M KCl/0.2 M NaCl) (Supplementary Fig. 5a-h). A summary of the $I-V$ relation for the MmPiezo1 channel, extracted from 204,088 events obtained in three experiments, shows that the single channel current is ohmic between -100 and 200 mV with a slope conductance of 102 ± 2 pS (Supplementary Fig. 5i). The current reversed direction at 0.0 ± 0.3 mV, demonstrating that the channel does not select between K^+ and Na^+ , and importantly, displays open channel block by ruthenium red (Supplementary Fig. 5j-l).

The difference in γ between overexpressed MmPiezo1 in cells and reconstituted MmPiezo1 in lipid bilayers may be attributed to many variables, including the distinct lipid environments which are known to influence conductance measurements strongly²⁸⁻³². Moreover, the ionic conditions used in the two systems are different, as divalent cations present in HEK293T cell-attached experiments also affect the conductance values. Indeed, when divalent cations are excluded from the recording pipette, γ of MmPiezo1-induced currents in HEK293T cells is $58.0 \text{ pS} \pm 1.5 \text{ pS}$ (150 mM NaCl solution, Supplementary Fig. 6), compared to 29.9 ± 1.5 pS in the presence of divalent ions (Fig. 3). The near equivalence of γ values together with the similar pattern of channel activity demonstrates that reconstitution

of MmPiezo1 into two distinct bilayer systems produces channels with identical functional properties (Supplementary Table 3).

Future reconstitution and recording of DmPiezo in lipid bilayers will show whether the difference in conductance between MmPiezo1 and DmPiezo arises from intrinsic properties. The membrane environment and lipid composition are known to modulate the activity of the embedded channel proteins in a drastic and deterministic manner (for example, see refs 28-32). It is not entirely surprising that the conditions to emulate the cellular environment in the reconstituted system in terms of the mechanical state of the membrane or its lipid composition have thus far been inadequate to retrieve the activation features of mechanically activated ion channels. Furthermore, the complexity of protein clusters and dynamic cytoskeletal interacting partners at the cell membrane³³ introduce regulatory constraints on channel activity. Further investigation may clarify whether Piezo ion channel subunits are intrinsically mechanosensitive or use unknown interacting partners to sense membrane tension.

Concluding remarks

We provide compelling evidence to support the hypothesis that Piezo proteins are indeed ion channels. First, overexpression of DmPiezo or MmPiezo1 in a human cell line gives rise to mechanically activated channels with distinct biophysical and pore-related properties. Second, isolated Piezo complexes do not contain detectable amounts

28. Ermakov, Y. A., Kamaraju, K., Sengupta, K. & Sukharev, S. Gadolinium ions block mechanosensitive channels by altering the packing and lateral pressure of anionic lipids. *Biophys. J.* **98**, 1018–1027 (2010).
29. Gambale, F. & Montal, M. Characterization of the channel properties of tetanus toxin in planar lipid bilayers. *Biophys. J.* **53**, 771–783 (1988).
30. Oliver, D. *et al.* Functional conversion between A-type and delayed rectifier K⁺ channels by membrane lipids. *Science* **304**, 265–270 (2004).
31. Schmidt, D. & MacKinnon, R. Voltage-dependent K⁺ channel gating and voltage sensor toxin sensitivity depend on the mechanical state of the lipid membrane. *Proc. Natl Acad. Sci. USA* **105**, 19276–19281 (2008).
32. Tao, X. & MacKinnon, R. Functional analysis of Kv1.2 and paddle chimera Kv channels in planar lipid bilayers. *J. Mol. Biol.* **382**, 24–33 (2008).
33. Hartman, N. C. & Groves, J. T. Signaling clusters in the cell membrane. *Curr. Opin. Cell Biol.* **23**, 370–376 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. H. Ulbrich for providing Ca²⁺ channel-, NMDA receptor- and CNG channel–GFP fusion constructs used as controls for photobleaching

experiments. This research was supported by grants from the National Institutes of Dental and Craniofacial Research, Neurological Disorders, General Medical Sciences, and by The Genomics Institute of the Novartis Research Foundation. B.X. and J.G. are postdoctoral fellowship recipients from the American Heart Association and the NIH, respectively.

Author Contributions B.C. performed and analysed electrophysiological experiments. B.X. performed and analysed biochemical experiments. J.S.S. and R.S. performed the reconstitution experiments and together with M.M. analysed the single channel data. J.G. and K.S.S. performed and analysed photo-bleaching experiments. S.E.K. cloned the *Dmpiezo* gene. M.S. initiated biochemical experiments. J.M. generated GFP–*MmPiezo1* and the mRNA used for oocyte injection. A.E.D. provided technical help for oocyte experiments. A.P., B.C., B.X., J.G., J.S.S., R.S., and M.M. wrote the manuscript.

Author Information The DmPiezo sequence has been deposited in GenBank under accession number JQ425255. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.P. (apatapou@gnf.org) and M.M. (mmontal@ucsd.edu).

METHODS

Cloning of *Drosophila piezo* full-length cDNA. The *Drosophila piezo* gene (GenBank accession number JQ425255) was cloned from adult *Drosophila* poly(A)⁺ RNAs (Clontech) by RT-PCR. Primers for RT-PCR were designed based on the annotated sequence of CG8486. Two fragments of 2 kb and 6.5 kb were amplified and cloned sequentially into pIRES2-EGFP expression vector. Each cloning step was sequence verified. Full-length *Drosophila piezo* gene is 8,355 bp in length. The protein sequence of DmPiezo is shown in Supplementary Fig. 1.

Cell culture and transient transfection. Human embryonic kidney 293T (HEK293T), NIH-3T3, F11 and HeLa cells were grown in Dulbecco's Modified Eagle Medium containing 4.5 mg ml⁻¹ glucose, 10% fetal bovine serum, 50 U ml⁻¹ penicillin and 50 µg ml⁻¹ streptomycin. Cells were plated onto poly-lysine-coated 12-mm round glass coverslips placed in 24-well plates and transfected using lipofectamine 2000 (Invitrogen) according to the manufacturer's instruction. 500–1,000 ng ml⁻¹ of plasmid DNA was transfected and cells were recorded 12–48 h later.

Electrophysiology. Patch-clamp experiments were performed in standard whole-cell or cell-attached recordings using an Axopatch 200B amplifier (Axon Instruments). Patch pipettes had resistance of 2–3 MΩ when filled with an internal solution consisting of (in mM) 133 CsCl, 10 HEPES, 5 EGTA, 1 CaCl₂, 1 MgCl₂, 4 MgATP and 0.4 Na₂GTP (pH adjusted to 7.3 with CsOH). The extracellular solution consisted of (in mM) 130 NaCl, 3 KCl, 1 MgCl₂, 10 HEPES, 2.5 CaCl₂, 10 glucose (pH adjusted to 7.3 with NaOH). For cell-attached recordings, pipettes were filled with a solution consisting of (in mM) 130 NaCl, 5 KCl, 10 HEPES, 1 CaCl₂, 1 MgCl₂, 10 TEA-Cl (pH 7.3 with NaOH), except for Supplementary Fig. 6 where the internal solution was (in mM) 150 NaCl, 10 HEPES (pH adjusted to 7.3 with NaOH). External solution used to zero the membrane potential consisted of (in mM) 140 KCl, 10 HEPES, 1 MgCl₂, 10 glucose (pH 7.3 with KOH). All experiments were done at room temperature. Currents were sampled at 50 or 20 kHz and filtered at 5 or 2 kHz. Voltages were not corrected for a liquid junction potential. Leak currents before mechanical stimulations were subtracted off-line from the current traces. 10 mM ruthenium red stock solution was prepared in water.

Mechanical stimulation. For whole-cell recordings mechanical stimulation was achieved using a fire-polished glass pipette (tip diameter 3–4 µm) positioned at an angle of 80° to the cell being recorded. Downward movement of the probe towards the cell was driven by a Clampex controlled piezo-electric crystal micro-stage (E625 LVPZT Controller/Amplifier; Physik Instrumente). The probe was typically positioned ~2 µm from the cell body. The probe had a velocity of 1 µm ms⁻¹ during the ramp segment of the command for forward motion and the stimulus was applied for 150 ms. To assess the mechanical sensitivity of a cell, a series of mechanical steps in 1 µm increments was applied every 10–20 s, which allowed full recovery of mechanosensitive currents. Inward mechanically activated currents were recorded at a holding potential of –80 mV. For *I*–*V* relationship recordings, voltage steps were applied 0.7 s before the mechanical stimulation from a holding potential of –60 mV.

For cell-attached recordings, membrane patches were stimulated with brief negative pressure pulses through the recording electrode using a Clampex controlled pressure clamp HSPC-1 device (ALA-scientific). Unless otherwise stated, stretch-activated channels were recorded at a holding potential of –80 mV with pressure steps from 0 to –60 mmHg (–10 mmHg increments). Current–pressure relationships were fitted with a Boltzmann equation of the form: $I(P) = [1 + \exp(-(P - P_{50})/s)]^{-1}$, where *I* is the peak of stretch-activated current at a given pressure, *P* is the applied patch pressure (in mmHg), *P*₅₀ is the pressure value that evoked a current value which is 50% of *I*_{max}, and *s* reflects the current sensitivity to pressure.

Single-channel amplitude characterization was performed on patches that showed strong stretch-activated current activity at –80 mV using increasing steps of negative pressure up to –60 mmHg. Similar activity was never present in control-transfected cells. Negative pressure steps were then reduced to low to moderate level (–5 to –20 mmHg) allowing detection of single channel openings.

Subunit counting. For oocyte injection, all construct plasmids were linearized at C terminus with NheI, HindIII or NotI and DNA transcribed with T7 mMessage mMachin Kit (Ambion) and poly(A)-tailing Kit (Ambion) and cleaned with LiCl precipitation. 50 nl of 0.2 µg µl⁻¹ mRNA was injected into *Xenopus* oocytes (Nasco).

For acquisition, 12–24 h after injection, oocytes were osmotically shocked in stripping buffer (in mM: 220 *N*-methyl glucamine aspartate, 10 HEPES, 1 MgCl₂) and mechanically de-vitellinated. MatTek dishes (MatTek Corporation) were prepared by sonication in 1 M KOH to remove background fluorescence and further sonicated in MilliQ dH₂O. Oocytes were placed onto MatTek dishes into SOS buffer (in mM: 100 NaCl, 2 KCl, 1.8 CaCl₂·H₂O, 1 MgCl₂·6H₂O, 5 HEPES,

2.5 Na pyruvate and 50 µg ml⁻¹ gentamicin, pH 7.0). The preparations were imaged on an inverted Nikon Ti-E fluorescence TIRF microscope (Nikon Corporation) and imaged with a high numerical aperture objective (Nikon 100× PlanApo, NA1.49) with an additional ×1.5 Optovar magnification. eGFP fusion proteins were excited with a 488-nm Coherent laser (Coherent, Inc.) and images were collected with an Andor iXon DU-897 EMCCD camera. Sixty-second movies were collected at 100-ms exposures, for a frame rate of 10 Hz.

Using Nikon Elements software, movies were duplicated and processed with a rolling average of 2. A second duplicate was filtered with a low-pass kernel of 7, to remove background. The low-pass images were subtracted from the averaged images, to produce the movies used for analysis. Non-overlapping 4 × 4 pixel regions of interest were drawn around randomly selected spots that were clearly separated from neighbouring bright pixels. The spots were required to fit entirely within the 4 × 4 pixel regions. Pixel size was 0.11 µm. The average intensity of each region was plotted over the length of the movies. Traces were discarded if the intensity increased after an initial decrease, if the fluorescent spot moved out of the region, or if the fluorescent signal showed a continuous decay instead of step-wise bleaching. Finally, the number of bleaching steps was counted for each spot.

MmPiezo1–GST purification. The MmPiezo1–GST construct was subcloned by inserting a GST encoding sequence from *Schistosoma japonicum* into the MmPiezo1 construct⁸ at the 3' end of MmPiezo1 cDNA sequence using the AscI and SacII restriction enzyme sites. The resulting MmPiezo1–GST fusion protein has 2,773 amino acids.

After incubation with cell lysates overnight at 4 °C, the glutathione beads were washed four times in a buffer containing 25 mM NaPIEPES, 140 mM NaCl, 0.6% CHAPS, 0.14% phosphatidylcholine (PC), 2.5 mM dithiothreitol (DTT), and a cocktail of protease inhibitors and eluted with 100 mM glutathione in a buffer containing 25 mM NaPIEPES, 50 mM Tris, 0.6% CHAPS, 0.14% PC, 2.5 mM DTT and a cocktail of protease inhibitors. The eluant was dialysed against a buffer containing 25 mM NaPIEPES, 0.6% CHAPS, 0.14% PC, 2.5 mM DTT and a cocktail of protease inhibitors. The purified samples were kept at 4 °C. Samples purified according to this protocol were used for all the biochemical work and the initial reconstitution experiments. However, because retrieval of channel activity from the reconstituted MmPiezo1–GST fluctuated from preparation to preparation, we adopted an alternative purification protocol involving the membrane fraction as the starting material. Specifically, 24 h after transfection, cells were collected and homogenized in a buffer containing 25 mM NaPIEPES, 50 mM NaCl, 2.5 mM DTT, and a cocktail of protease inhibitors. The cell suspension was forced to go through a 25.5 G needle for 20 times and centrifuged at 1,000g for 15 min at 4 °C. The supernatant was collected and centrifuged at 167,000g for 30 min at 4 °C. The resulting membrane fraction was washed three times (using a buffer containing 25 mM NaPIEPES, 150 mM NaCl, 2.5 mM DTT, and a cocktail of protease inhibitors) and used as the starting material for MmPiezo1–GST purification using the same procedure described above. Purification from the membrane fraction greatly reduced the content of endogenous GST proteins and significantly enhanced the frequency of retrieval of MmPiezo1 channel activity after reconstitution (Fig. 5, Supplementary Fig. 5 and Supplementary Table 3).

NativePAGE Novex Bis-Tris gel. The purified MmPiezo1–GST proteins and control samples were subjected to 3–12% NativePAGE Novex Bis-Tris gel for native (non-denaturing) electrophoresis according to the User Manual (Invitrogen). In brief, samples were mixed with NativePAGE Sample Buffer and NativePAGE 5% G-250 Sample Additive and then subjected to electrophoresis at 150 V for 2 h. The use of G-250 charge-shift in NativePAGE gels results in protein resolution based upon protein size and therefore allows accurate size estimation of native protein complexes³⁴. However, the native protein conformation may give an expected size estimation error of ~15%. After electrophoresis, the native gel was then either visualized by a fast Coomassie G-250 staining or transferred to a PVDF membrane for western blotting with an antibody specifically against Piezo1 proteins.

Formaldehyde and paraformaldehyde crosslinking. The purified MmPiezo1–GST proteins were treated with or without 0.1% formaldehyde at room temperature for different periods of time and then mixed with NuPAGE LDS Sample Buffer and NuPAGE Reducing Agent (Invitrogen), followed by heating at 70 °C for 10 min to denature the protein. The treated samples were subjected to 3–8% NuPAGE Tris-Acetate gel electrophoresis under denaturing conditions. For live cell crosslinking, 0.25% concentration of PFA was added to the cell culture medium and kept at room temperature for 10 min, followed by adding 125 mM glycine to stop the PFA crosslinking reaction. Treated cells were collected and subjected to sequential steps of protein purification, 3–8% NuPAGE Tris-Acetate gel electrophoresis under denaturing conditions or 3–12% NativePAGE Novex Bis-Tris gel for native (non-denaturing) electrophoresis, and western blotting with the anti-Piezo1 antibody.

Western blotting. After electrophoresis, either the native or denaturing PAGE gels were transferred to PVDF membranes. Transferring protein from native gel to PVDF membranes was conducted according to instructions for NativePAGE Novex Bis-Tris gel system. Transferred PVDF membranes were blocked with 5% milk in TBS buffer with 0.1% Tween-20 (TBST buffer) at room temperature for 1 h, and then incubated with the anti-Piezo1 antibody (1:200) at 4 °C overnight. The membranes were washed with TBST buffer and incubated with peroxidase-conjugated anti-rabbit IgG secondary antibody (1:10,000) at room temperature for 1 h. Proteins were detected with the ECL plus detection kit (GE Healthcare).

Mass spectrometry. Purified samples were separated on the 3–12% NativePAGE Novex Bis-Tris gel and visualized by fast Coomassie G-250 staining. The gel band containing the MmPiezo1–GST complex or the corresponding blank band from the control sample near the 1,236 kDa molecular marker was excised and subjected to the Scripps Center for Mass Spectrometry for analysis. In brief, the gel bands were destained, reduced with 10 mM DTT, alkylated with 55 mM iodoacetamide, and digested with Trypsin overnight before analysis using the nano-LC-MS/MS. The nano-LC-MS/MS data obtained on a LTQ ion trap mass spectrometer was searched using the MmPiezo1–GST protein sequence and NCBI *Homo sapiens* database. In separate sets of experiments, the purified MmPiezo1–GST and control solution samples before gel electrophoresis were subjected to mass spectrometry (Supplementary Table 2).

Reconstitution into proteoliposomes or DIBs, single channel recordings and analysis. Purified MmPiezo1–GST protein was reconstituted into asolectin (soybean polar lipid extract, Avanti) liposomes (10 mg ml^{-1}) by incubating the mixture (lipid/protein mass ratios between 2,000:1 and 1,000:1; this corresponds to a molar lipid/protein ratio of $\sim 800,000\text{--}400,000:1$) on ice for 5 min followed by $\times 20$ dilution in 200 mM KCl, 5 mM MOPS pH 7.0 and incubated with rotation at room temperature for 20 min. Biobeads were added to mixture and incubated with rotation for 1 h. Thereafter, biobeads were removed by filtration and a new batch of beads was added. After 30 min incubation, the biobeads were filtered and the sample was centrifuged at 60,000 r.p.m. for 60 min at 8 °C. The proteoliposome pellet was re-suspended in 40 μl of the same buffer and used to place two 25 μl drops on a cover slide. The samples were dried under vacuum for $>16 \text{ h}$ at 4 °C. Samples were hydrated with 25 μl of the same buffer and allowed to sit for 2 h before starting recordings. Thereafter, 2–3 μl of proteoliposomes were withdrawn from the edge of the spots on the cover slide and transferred to the recording chamber. After 5 min, the chamber was slowly filled with recording solution. Multi-G Ω seals were made to proteoliposomes immobilized at the bottom of the

recording chamber. At that time, the proteoliposome patch was excised and brought through the air–water interface. Excised patches were used³⁵. Pipette and bath solution contained (in mM) 200 KCl, 5 MOPS titrated to pH 7.0 with KOH. Capillaries of borosilicate glass from Sigma were pulled to yield resistances of 1–2 M Ω when immersed in recording solution.

Droplet interface lipid bilayers (DIBs) were formed between two lipid monolayer-encased aqueous nanolitre droplets submerged in hexadecane²³. Liposomes were composed of 1,2-diphytanoyl-*sn*-glycero-3 phosphocholine (DPhPC) or 90% (mole/mole) DPhPC and 10% of the negatively charged lipid, 1,2-dioleoyl-*sn*-glycero-3-phosphatidic acid (DOPA) (Avanti Polar Lipids). MmPiezo1 was diluted directly into the liposome suspension to yield a final concentration of $2\text{--}5 \text{ ng ml}^{-1}$. The electrode carrying the droplet with MmPiezo1 and desired buffer–lipid mix (in mM, 500 KCl, 10 HEPES, pH 7.4, 0.5 lipid solution of DPhPC) was connected to the grounded end of the amplifier head-stage (Axopatch 200B). The second electrode, in a droplet containing the same buffer and 10% DOPA:90% DPhPC, was connected to the working end of the head-stage. Where indicated, ruthenium red or 0.2 M NaCl was injected using a nano-injector (WPI, Inc.).

For proteoliposome patches, records were acquired at a sampling frequency of 40 kHz and filtered online to 5 kHz with a 3-pole Bessel filter before digitization; for DIBs, data acquisition was at 10 kHz and filtered at 2 kHz. For analysis and presentation, records were filtered to 1 kHz with a low-pass Gaussian filter. Transitions were detected by the half-threshold method implemented in Clampfit (proteoliposomes) and by the segmental *k*-means method (SKM) implemented in QuB (DIBs). Transitions $\leq 0.5 \text{ ms}$ were excluded from the pool for analysis to correct for detection of false and missed events. Data were analysed using Clampfit v.9.2 software (Axon Instruments), QuB, Excel 2007 (Microsoft), and Igor Pro (Wavemetrics). γ was calculated from Gaussian fits to currents histograms. All statistical values represent mean \pm s.e.m., unless otherwise indicated. *n* and *N* denote number of experiments and number of events, respectively. All experiments were done at room temperature.

34. Schagger, H., Cramer, W. A. & von Jagow, G. Analysis of molecular masses and oligomeric states of protein complexes by blue native electrophoresis and isolation of membrane protein complexes by two-dimensional native electrophoresis. *Anal. Biochem.* **217**, 220–230 (1994).
35. Gambale, F. & Montal, M. Voltage-gated sodium channels expressed in the human cerebellar medulloblastoma cell line TE671. *Brain Res. Mol. Brain Res.* **7**, 123–129 (1990).

The sirtuin SIRT6 regulates lifespan in male mice

Yariv Kanfi¹, Shoshana Naiman^{1*}, Gail Amir^{2*}, Victoria Peshti¹, Guy Zinman³, Liat Nahum¹, Ziv Bar-Joseph³ & Haim Y. Cohen¹

The significant increase in human lifespan during the past century confronts us with great medical challenges. To meet these challenges, the mechanisms that determine healthy ageing must be understood and controlled. Sirtuins are highly conserved deacetylases that have been shown to regulate lifespan in yeast, nematodes and fruitflies¹. However, the role of sirtuins in regulating worm and fly lifespan has recently become controversial². Moreover, the role of the seven mammalian sirtuins, SIRT1 to SIRT7 (homologues of the yeast sirtuin Sir2), in regulating lifespan is unclear³. Here we show that male, but not female, transgenic mice overexpressing *Sirt6* (ref. 4) have a significantly longer lifespan than wild-type mice. Gene expression analysis revealed significant differences between male *Sirt6*-transgenic mice and male wild-type mice: transgenic males displayed lower serum levels of insulin-like growth factor 1 (IGF1), higher levels of IGF-binding protein 1 and altered phosphorylation levels of major components of IGF1 signalling, a key pathway in the regulation of lifespan⁵. This study shows the regulation of mammalian lifespan by a sirtuin family member and has important therapeutic implications for age-related diseases.

Sirtuins are highly conserved NAD⁺-dependent deacetylases that have been shown to regulate lifespan in several organisms. Increasing the sirtuin level through genetic manipulation extends the lifespan of yeast, nematodes and flies¹. Yet, despite many publications supporting a pro-longevity role for sirtuins, there has been recent debate about the direct role of *Caenorhabditis elegans* and *Drosophila melanogaster* SIR-2 in ageing and lifespan extension in response to calorie restriction (also known as dietary restriction)^{2,6,7}. Some mammalian sirtuins have been shown to regulate age-related diseases, but mice that overexpress SIRT1 have the same lifespan as control, wild-type (WT), mice⁸. Thus, the role of SIRT1 and other mammalian sirtuins in regulating mammalian lifespan is unclear³.

Several key findings support a potential role for SIRT6 in regulating mammalian lifespan. SIRT6-deficient mice are small and have severe metabolic defects, and by 2–3 weeks of age, they develop abnormalities that are usually associated with ageing⁹. In addition, SIRT6 regulates nuclear factor- κ B signalling, which controls ageing-associated changes in gene expression¹⁰. Recently, we showed that SIRT6 levels increase in rats that are fed a calorie-restricted diet¹¹, and transgenic mice that overexpress exogenous mouse SIRT6 (*Sirt6*-transgenic mice; also known as MOSES mice)⁴ are protected against the physiological damage caused by diet-induced obesity, including triglyceride and low-density-lipoprotein-associated cholesterol accumulation in the serum, increased body fat and reduced glucose tolerance. In normal animals, these metabolic defects become apparent by middle age, whereas their appearance is delayed in animals fed a calorie-restricted diet. Thus, in this study we sought to determine whether *Sirt6*-transgenic mice remain healthy for longer and have a longer lifespan than wild-type mice.

The lifespan of *Sirt6*-transgenic mice was examined in comparison to their control littermates. *Sirt6*-transgenic mice were produced on a segregating stock containing equal contributions from C57BL/6J and BALB/cOlaHsd mouse strains, both of which are considered to be long

lived¹². The study was carried out on 245 mice (119 males and 126 females) from two transgenic lines (line 55 and line 108) generated from two separate founders. Log-rank test analysis showed significant differences in the survival curves between male WT and male transgenic mice, but not between female WT and female transgenic mice, for both lines (Fig. 1a–d and Supplementary Table 1). Relative to male WT littermates, the median lifespan of male *Sirt6*-transgenic mice increased by 14.5% and 9.9%, and the mean lifespan increased by 14.8% and 16.9%, for line 55 and 108, respectively (log-rank test, $\chi^2 = 10.529$, d.f. = 1 and $P = 0.001$ for line 55; and $\chi^2 = 4.225$, d.f. = 1 and $P = 0.040$ for line 108). In female *Sirt6*-transgenic mice, no significant increase in median or mean lifespan was found relative to female WT littermates for either line (log-rank test, $\chi^2 = 0.009$,

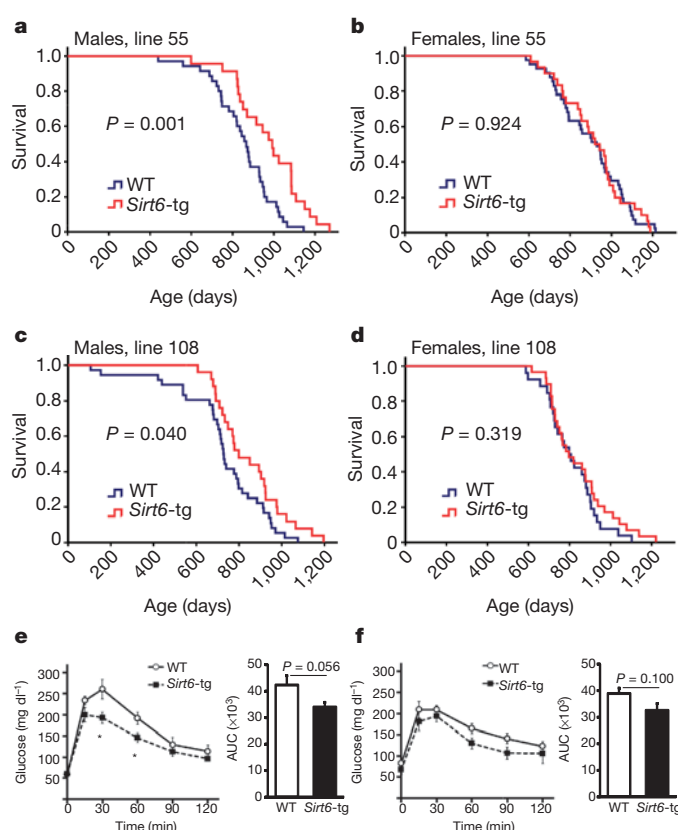


Figure 1 | Extended lifespan of male *Sirt6*-transgenic mice. Kaplan-Meier survival curves for male and female WT and *Sirt6*-transgenic (*Sirt6*-tg) mice from two transgenic lines, line 55 (a, b) and line 108 (c, d). P values were derived from log-rank calculations. Glucose tolerance testing was carried out in WT and *Sirt6*-transgenic males (e) and females (f) at 19 months (572–577 days) of age (males, $n = 6$ per genotype; females, $n = 4$ per genotype). The area under the curve (AUC) for each glucose tolerance test is shown on the right (e, f; y axis values shown are the AUC divided by 1,000). The values shown are mean \pm s.e.m. *, $P < 0.05$ (two-tailed t -test).

¹The Mina & Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan 52900, Israel. ²Department of Pathology, Hadassah Medical Center and Hebrew University, Kiryat Hadassah, Jerusalem 91120, Israel. ³Lane Center for Computational Biology, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania 15217, USA.

*These authors contributed equally to this work.

d.f. = 1 and $P = 0.924$ for line 55; and $\chi^2 = 0.993$, d.f. = 1 and $P = 0.319$ for line 108). Relative to WT littermates, the maximum lifespan of transgenic males (that is, the mean lifespan of the oldest 10% of a cohort to die) increased by 15.8% and 13.1% for line 55 and 108, respectively. Comparison of the maximum lifespan of WT and *Sirt6*-transgenic mice using the quantile regression approach at the ninetieth percentile¹³ showed a significant difference between males in one line only ($P = 0.03$ and $P = 0.11$ for line 55 and 108, respectively) and no difference for females ($P = 0.45$ and $P = 0.67$ for line 55 and 108, respectively). Cox regression analysis (using the stepwise backward, Wald method) with the recruitment date, parental identity, gender, genotype and mouse line as main effects and line-by-genotype as the interaction variable showed an additive effect of genotype and line (Supplementary Table 2). However, there was no interaction between mouse line and genotype ($P = 0.693$), indicating that SIRT6 overexpression had an equivalent effect on the mortality of both lines. In summary, our data show that SIRT6 overexpression increased the longevity of males but not females.

SIRT6 has been shown to regulate genomic stability and metabolism^{4,9}, two important contributors to longevity. Loss of genomic stability is known to be an important aspect of cancer. Post-mortem gross and microscopic examination of the WT and transgenic mice revealed malignant tumours in a variety of organs, with the highest incidence of tumours in all mice being in the lungs. No significant differences in tumour spectrum or incidence were found between WT and transgenic mice (Supplementary Table 3). Similarly, pathological analysis revealed no differences between WT and transgenic males in the incidence of non-neoplastic findings (for example, diffuse mesangial sclerosis and pulmonary emphysema) or age-related pathologies (for example, femoral osteoporosis, basal ganglia calcification and adrenal cortical hyperplasia) (data not shown). Interestingly, the median lifespan of *Sirt6*-transgenic mice with lung tumours showed a trend towards being longer (by 11.7%) than that of WT mice with lung tumours. Therefore, the hypothesis that the effect of SIRT6 on lung cancer has a role in SIRT6's pro-longevity effect cannot be entirely excluded. However, given the proportion of mice with lung tumours in each genotype, a protective role of SIRT6 against lung cancer is likely to contribute only partially to the pro-longevity effect (Supplementary Information). Thus, further studies are required to evaluate the contribution of SIRT6 to age-sensitive traits, in addition to its effect on lung cancer.

The protective role of SIRT6 against metabolic disorders that are induced by a high-fat diet⁴ suggests that SIRT6 might positively affect age-associated metabolic disorders, such as declining insulin sensitivity and impaired glucose tolerance. No significant differences in glucose metabolism were found between young (4–7 month old) WT and *Sirt6*-transgenic mice (data not shown). However, an intraperitoneal glucose tolerance test showed that old *Sirt6*-transgenic mice (19 months old, the maximum age of WT mice before a considerable proportion of the litter died) displayed a trend towards improved glucose homeostasis compared with WT mice of the same age (Fig. 1e, f). An analysis of variance (ANOVA) test for the area under the curve (AUC) values of the glucose tolerance tests indicated no sex-specific effect but showed a significant effect of genotype ($P = 0.016$). Therefore, although SIRT6 overexpression had a positive effect on glucose homeostasis in old mice, this finding cannot explain the sexual dimorphism in longevity.

To understand further the mechanisms of the gender-specific lifespan extension in *Sirt6*-transgenic mice, we used whole genome microarray analysis to examine differential gene expression in the livers of animals of both sexes (Supplementary Table 4). In agreement with the sexual dimorphism in liver gene expression¹⁴, differential expression analysis using Significance Analysis of Microarrays (SAM) software¹⁵ showed that the most extensive gene expression differences occurred between genders (Supplementary Table 5). Notably, significant differences were also found between *Sirt6*-transgenic and WT males, but the differences between *Sirt6*-transgenic and WT females were minor

(Supplementary Table 5). ANOVA analysis uncovered a subset of genes whose expression differed significantly between genotypes and that were gender-specific (Supplementary Table 5). Gene Ontology (GO) functional analysis showed that the differentially expressed gene set between *Sirt6*-transgenic males and WT males is significantly enriched for categories related to metabolism and cellular responses (Supplementary Table 6). We next compared this differentially expressed gene set with the set of genes that was differentially expressed between male and female WT mice. This analysis revealed a significant similarity between the two gene sets. Of the differentially expressed genes in *Sirt6*-transgenic males, 50% (41 of 82) were also differentially expressed between male and female WT mice ($P = 0$) (Fig. 2a, b and Supplementary Table 5).

To confirm the microarray results, 11 of the differentially expressed genes in male *Sirt6*-transgenic mice were selected for validation by quantitative PCR. The expression pattern of all of these 11 genes confirmed the microarray data (Figs 2c and 3c). Moreover, to examine whether the transcriptional changes due to SIRT6 are mouse-line-specific, the expression of several of these genes was followed in another transgenic line, and the same pattern of transcriptional changes was observed (Supplementary Fig. 1). Calorie restriction and starvation^{16–19} have previously been shown to have a similar effect to SIRT6 overexpression on the transcription of several genes (30% of the differentially expressed genes in male *Sirt6*-transgenic mice showed a similar expression pattern in male mice fed a calorie-restricted diet¹⁶). For

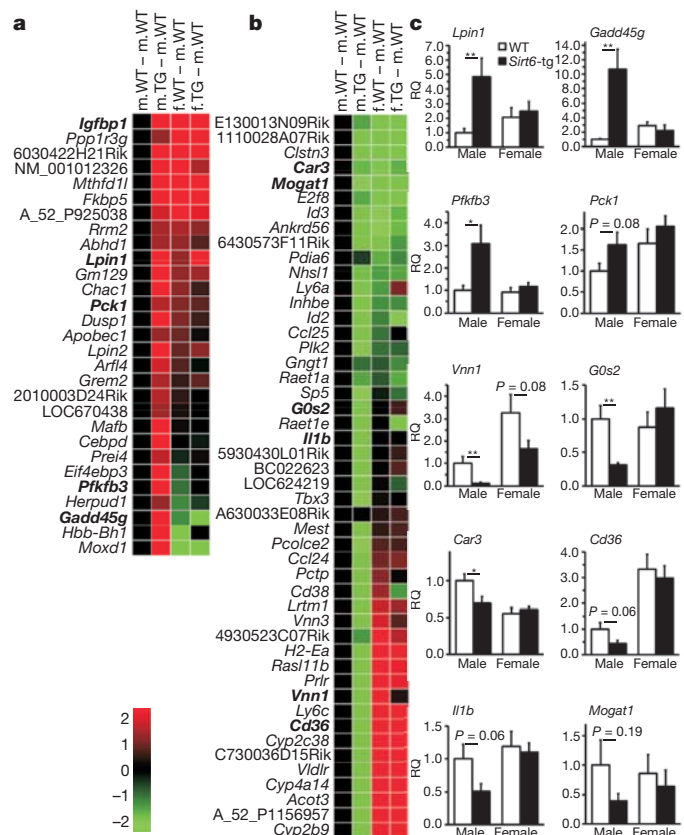


Figure 2 | Expression profile of differentially expressed genes in male *Sirt6*-transgenic mice. a, b, Heat maps displaying the significantly upregulated (red) and downregulated (green) genes in *Sirt6*-transgenic males (m.TG) compared with WT males (m.WT). The expression profile of these genes in WT females (f.WT) or *Sirt6*-transgenic females (f.TG) compared with WT males is also illustrated. Statistical analysis was performed using all 24 arrays. The quantitative-PCR-validated genes are shown in bold. c, The relative expression levels of hepatic genes were confirmed by quantitative PCR in 20 male and 20 female mice. The values shown are mean \pm s.e.m. *, $P < 0.05$; **, $P < 0.01$; n.s. = not significant. RQ, relative quantification.

example, the upregulated genes *Lpin1*, *Lpin2*, *Gadd45g*, *Fkbp5*, *Dusp1* and *Cebpd* and the downregulated genes *Vnn1*, *Vnn3*, *Pctp*, *Vldlr*, *Car3* and *G0s2* in the expression profile of male *Sirt6*-transgenic mice are also differentially expressed in the livers of mice fed a calorie-restricted diet^{16–18}.

A key factor in the regulation of lifespan is the IGF1 signalling pathway. Worms and flies with a mutated insulin/IGF1 receptor and mice that are heterozygous for the IGF1 receptor have an extended lifespan⁵. Moreover, rodents fed a calorie-restricted diet have lower IGF1 levels early in life than rodents fed a normal chow diet, and many rodent genetic models with a prolonged lifespan have lower levels of serum IGF1 or IGF1 signalling than do control groups^{5,20}. Although no difference was found between WT and *Sirt6*-transgenic females, young transgenic males (6 months old) had lower serum IGF1 levels than WT male littermates (Fig. 3a), and these IGF1 levels in *Sirt6*-transgenic males were similar to those in all females. This significant difference in IGF1 levels between young transgenic and WT males was sustained until 19 months of age (Fig. 3b). In line with this finding, one of the genes that was highly upregulated in *Sirt6*-transgenic males, to the same levels as in WT or *Sirt6*-transgenic females, was the gene encoding IGF-binding protein 1 (IGFBP1) (Fig. 3c). IGFBP1 is thought to be the main short-term modulator of IGF1 bioavailability²¹. Calorie restriction increases the expression of IGFBP1 (ref. 17), and high levels of IGFBP1 correlate with protection against metabolic disorders²². No change was found in the expression of gene encoding

other IGF1-binding proteins, such as IGFBP3 and acid-labile subunit (ALS; also known as IGFBP3) (Fig. 3c).

To follow the changes in IGF1 signalling, components of this pathway were analysed in the three main metabolic tissues: liver, white adipose tissue (WAT) and muscle. Analyses included the phosphorylation levels of AKT activation sites (Thr 308 and Ser 473), FOXO1 (Thr 24) and FOXO3 (Thr 32). The most significantly decreased phosphorylation levels were observed in the perigonadal WAT of *Sirt6*-transgenic males in comparison to WT males (Fig. 3d–g and Supplementary Fig. 2a–d). The levels of phosphorylated AKT (on both activation sites), FOXO1 and FOXO3 in WAT were lower in the transgenic mice (Fig. 3e, f). Therefore, we further explored this pathway in WAT and found that the phosphorylation levels of the IGF1 receptor (Tyr 1135) and S6 (Ser 235/236) were lower in *Sirt6*-transgenic males than in the WT male littermates (Fig. 3d, g). Importantly, no significant change in the phosphorylation levels of these markers was observed in female mice (Fig. 3d–g and Supplementary Fig. 2a–d). Moreover, the decrease in the phosphorylation levels of AKT and FOXO proteins in male *Sirt6*-transgenic mice is in agreement with previous reports that show that lifespan is positively regulated by changes in IGF1 signalling in the whole organism, or specifically in the fat tissues, of nematodes and fruitflies^{5,23}.

There is much doubt about whether mammalian sirtuins regulate lifespan^{3,5,8}. Moreover, in the fly and nematode, a recent study challenged the role of sirtuins in regulating lifespan, claiming that the increased

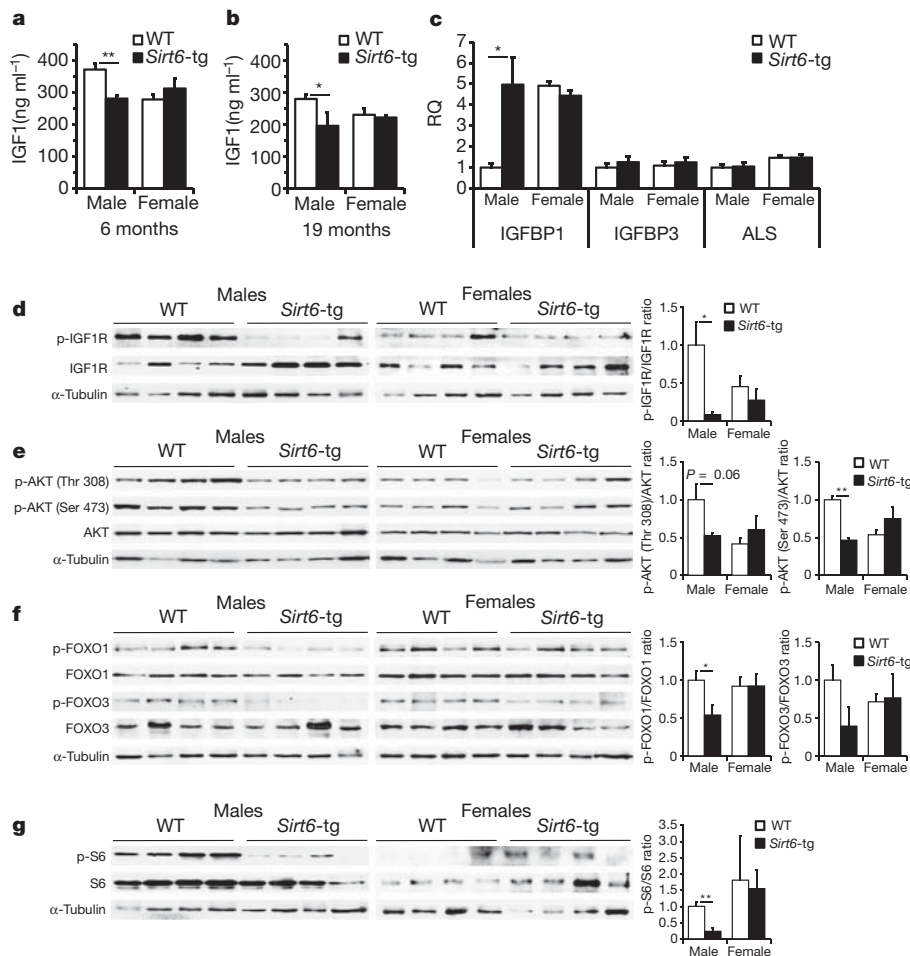


Figure 3 | Alterations in the IGF1-AKT pathway in *Sirt6*-transgenic males. **a, b**, Serum IGF1 levels in male and female WT and *Sirt6*-transgenic mice at 6 months (**a**) and 19 months (**b**) of age ($n = 4-7$). **c**, The relative expression of hepatic *Igf1*, *Igf1* and *Als* measured by quantitative PCR ($n = 4-7$). **d-g**, The phosphorylation levels of the IGF1 receptor (IGF1R) at Tyr 1135 (**d**), AKT at both the Thr 308 and Ser 473 activation sites (**e**), FOXO1 at Thr 24

and FOXO3 at Thr 32 (**f**), and S6 at Ser 235/236 (**g**) in perigonadal WAT ($n = 4$ mice per genotype). All mice were killed at the same time of day. The phosphorylated to unphosphorylated protein ratios, as determined by densitometry, are shown on the right. **a-g**, The values shown are the mean \pm s.e.m. *, $P < 0.05$; **, $P < 0.01$.

longevity observed in strains with SIR-2 overexpression is caused by differences in genetic background or by mutagenic effects of transgene insertion². To address potential complications owing to strain-specific effects and integration sites, we used a segregating background with equal contributions from the C57BL/6J and BALB/cOlaHsd mouse strains and studied two separate lines. Indeed, we showed that SIRT6 extends male lifespan regardless of the integration site (Supplementary Fig. 3) and in two control lines with different lifespans. Here, we reveal a role for the mammalian sirtuin SIRT6 in regulating lifespan. SIRT6 overexpression extends lifespan only in males, potentially by reducing IGF1 signalling specifically in WAT. Mice with a fat-specific insulin receptor gene knockout have been shown to have an increased mean lifespan of similar magnitude to the male transgenic mice in our study²³, demonstrating the central role of fat in regulating lifespan. Most genetic modifications of the IGF1 or insulin signalling pathway affect the lifespan of both genders or show a stronger effect in females. Yet here the effect of SIRT6 on IGF1 signalling was male specific. Therefore, further research is required to determine whether the effects of SIRT6 are blocked in females rather than enhanced in males. Taken together, our findings suggest that SIRT6 is an important regulator of mammalian longevity and indicate the feasibility of manipulating SIRT6 levels to treat age-related diseases.

METHODS SUMMARY

Sirt6-transgenic mice on the CB6F1 background, containing equal contributions from C57BL/6J and BALB/cOlaHsd mouse strains, were generated as described previously⁴, and the glucose tolerance tests and lifespan analyses were performed as described previously^{24–26}. Tissues were taken after natural death, fixed in formaldehyde for histopathological analysis, embedded in paraffin, sectioned, and stained with haematoxylin and eosin. Quantitative PCR was performed using Absolute Blue SYBR Green on a StepOnePlus instrument. Microarray sample labelling and hybridization were performed as previously described⁴, and data were normalized using the program dChip. Differentially expressed genes were identified using SAM and defined as those with a *q* value of <10.0% and a minimum of a 1.5 fold change.

Received 17 December 2010; accepted 19 December 2011.

Published online 22 February 2012.

1. Michan, S. & Sinclair, D. Sirtuins in mammals: insights into their biological function. *Biochem. J.* **404**, 1–13 (2007).
2. Burnett, C. *et al.* Absence of effects of Sir2 overexpression on lifespan in *C. elegans* and *Drosophila*. *Nature* **477**, 482–485 (2011).
3. Baur, J. A. *et al.* Dietary restriction: standing up for sirtuins. *Science* **329**, 1012–1014 (2010).
4. Kanfi, Y. *et al.* SIRT6 protects against pathological damage caused by diet-induced obesity. *Aging Cell* **9**, 162–173 (2010).
5. Kenyon, C. J. The genetics of ageing. *Nature* **464**, 504–512 (2010).
6. Viswanathan, M. & Guarente, L. Regulation of *Caenorhabditis elegans* lifespan by *sir-2.1* transgenes. *Nature* **477**, E1–E2 (2011).
7. Rizki, G. *et al.* The evolutionarily conserved longevity determinants HCF-1 and SIR-2.1/SIRT1 collaborate to regulate DAF-16/FOXO. *PLoS Genet.* **7**, e1002235 (2011).
8. Herranz, D. *et al.* Sirt1 improves healthy ageing and protects from metabolic syndrome-associated cancer. *Nature Commun.* **1**, 3 (2010).

9. Mostoslavsky, R. *et al.* Genomic instability and aging-like phenotype in the absence of mammalian SIRT6. *Cell* **124**, 315–329 (2006).
10. Kawahara, T. L. *et al.* SIRT6 links histone H3 lysine 9 deacetylation to NF- κ B-dependent gene expression and organismal life span. *Cell* **136**, 62–74 (2009).
11. Kanfi, Y. *et al.* Regulation of SIRT6 protein levels by nutrient availability. *FEBS Lett.* **582**, 2417–2423 (2008).
12. Yuan, R. *et al.* Aging in inbred strains of mice: study design and interim report on median lifespans and circulating IGF1 levels. *Aging Cell* **8**, 277–287 (2009).
13. Wang, C., Li, Q., Redden, D. T., Weindruch, R. & Allison, D. B. Statistical methods for testing effects on 'maximum lifespan'. *Mech. Ageing Dev.* **125**, 629–632 (2004).
14. Yang, X. *et al.* Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res.* **16**, 995–1004 (2006).
15. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98**, 5116–5121 (2001).
16. Estep, P. W. III, Warner, J. B. & Bulyk, M. L. Short-term calorie restriction in male mice feminizes gene expression and alters key regulators of conserved aging regulatory pathways. *PLoS ONE* **4**, e5242 (2009).
17. Swindell, W. R. Genes and gene expression modules associated with caloric restriction and aging in the laboratory mouse. *BMC Genomics* **10**, 585 (2009).
18. Selman, C. *et al.* Coordinated multitissue transcriptional and plasma metabolomic profiles following acute caloric restriction in mice. *Physiol. Genomics* **27**, 187–200 (2006).
19. Bauer, M. *et al.* Starvation response in mouse liver shows strong correlation with life-span-prolonging processes. *Physiol. Genomics* **17**, 230–244 (2004).
20. Holzenberger, M. *et al.* IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. *Nature* **421**, 182–187 (2003).
21. Lee, P. D., Giudice, L. C., Conover, C. A. & Powell, D. R. Insulin-like growth factor binding protein-1: recent findings and new directions. *Proc. Soc. Exp. Biol. Med.* **216**, 319–357 (1997).
22. Yeap, B. B. *et al.* IGF1 and its binding proteins 3 and 1 are differentially associated with metabolic syndrome in older men. *Eur. J. Endocrinol.* **162**, 249–257 (2010).
23. Bluhm, M., Kahn, B. B. & Kahn, C. R. Extended longevity in mice lacking the insulin receptor in adipose tissue. *Science* **299**, 572–574 (2003).
24. Harrison, D. E. *et al.* Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. *Nature* **460**, 392–395 (2009).
25. Conti, B. *et al.* Transgenic mice with a reduced core body temperature have an increased life span. *Science* **314**, 825–828 (2006).
26. Selman, C. *et al.* Ribosomal protein S6 kinase 1 signaling regulates mammalian life span. *Science* **326**, 140–144 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. S. Levy-Drummer, C. Wachtel, S. Schwarzbach and members of the Cohen laboratory for their comments on the manuscript. This study was supported by National Institutes of Health grant 1R01 GM085022 to Z.B.-J. and by grants from the Israeli Academy of Sciences, the United States - Israel Binational Science Foundation, the Israel Cancer Association, the Koret Foundation, the Israel Cancer Research Fund, the Israel Health Ministry, I-CORE program (41/1), the Israel Science Foundation and the European Research Council to H.Y.C.

Author Contributions H.Y.C. designed experiments, analysed data and contributed to writing the paper. Y.K. designed and performed experiments, analysed data and contributed to writing the paper. S.N. designed and performed experiments and contributed to writing the paper. G.A. performed the histopathological analysis. V.P. and L.N. performed experiments. G.Z. and Z.B.-J. developed analytical tools, analysed data and contributed to writing the paper. S.N. and G.A. contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to H.Y.C. (Haim.Cohen@biu.ac.il).

Imaging ultrafast molecular dynamics with laser-induced electron diffraction

Cosmin I. Blaga¹, Junliang Xu², Anthony D. DiChiara¹, Emily Sistrunk¹, Kaikai Zhang¹, Pierre Agostini¹, Terry A. Miller³, Louis F. DiMauro¹ & C. D. Lin²

Establishing the structure of molecules and solids has always had an essential role in physics, chemistry and biology. The methods of choice are X-ray and electron diffraction, which are routinely used to determine atomic positions with sub-ångström spatial resolution. Although both methods are currently limited to probing dynamics on timescales longer than a picosecond, the recent development of femtosecond sources of X-ray pulses and electron beams suggests that they might soon be capable of taking ultrafast snapshots of biological molecules^{1,2} and condensed-phase systems^{3–6} undergoing structural changes. The past decade has also witnessed the emergence of an alternative imaging approach based on laser-ionized bursts of coherent electron wave packets that self-interrogate the parent molecular structure^{7–11}. Here we show that this phenomenon can indeed be exploited for laser-induced electron diffraction¹⁰ (LIED), to image molecular structures with sub-ångström precision and exposure times of a few femtoseconds. We apply the method to oxygen and nitrogen molecules, which on strong-field ionization at three mid-infrared wavelengths (1.7, 2.0 and 2.3 μm) emit photoelectrons with a momentum distribution from which we extract diffraction patterns. The long wavelength is essential for achieving atomic-scale spatial resolution, and the wavelength variation is equivalent to taking snapshots at different times. We show that the method has the sensitivity to measure a 0.1 Å displacement in the oxygen bond length occurring in a time interval of ~ 5 fs, which establishes LIED as a promising approach for the imaging of gas-phase molecules with unprecedented spatio-temporal resolution.

In conventional electron diffraction (CED), an external multi-kilovolt beam impinges on a molecular gas sample^{12–14}, producing a diffraction pattern by means of elastic scattering. The electron's high energy promotes core-penetrating collisions (impact parameters ≤ 0.1 Å), in which the electron–molecule interaction is dominated by the strong short-range atomic-like potential while the bonding valence electrons look transparent. Conceptually, LIED is a time-resolved equivalent of CED. The field-ionized and accelerated coherent electron wave packet rescattering from its parent molecular core¹⁰ (Fig. 1a) mimics the CED electron beam. If the rescattering wave packet is sufficiently energetic, the well-established bond length retrieval method of CED is applicable. We ensure that this is the case by using intense mid-infrared lasers for producing high-energy electron wave packets^{15,16} (see Methods). We can then exploit theory that establishes that large-momentum-transfer backscattered electrons in LIED are equivalent in spatial resolution to those in CED¹⁷, and use the LIED momentum angular distributions obtained at different wavelengths to observe changes in molecular bond lengths over an interval of a few femtoseconds.

The principle underpinning our approach is the strong-field rescattering model^{18,19} (Fig. 1a), in which a molecule is first ionized by an intense low-frequency field that promotes an electronic wave packet into the continuum at time t_b . Approximately half of the electrons, dubbed

direct (black dashed line in Fig. 1a), drift away from the vicinity of the molecule and are detected as low-momentum events. At a different t_b , the electron is accelerated by the laser field but follows a trajectory (magenta dashed line) that returns it to the core, where it can rescatter at time t_r . During the propagation time, $\Delta t = t_r - t_b$, the electron acquires significant kinetic energy from the field before rescattering. As the deBroglie wavelength of the returning wave packet becomes comparable to the molecular dimension, the interatomic spacing can be determined. At t_r the wave packet can either recombine to emit a high-energy photon^{20,21} (high harmonic generation) or scatter, producing a far-field diffraction pattern. Both processes generate signals that convey structural information about the parent or target molecule, but only LIED provides a direct conceptual link to the CED method.

In contrast with CED's field-free scattering, LIED is a strong-field process; the laser's influence therefore needs to be removed to allow the extraction of field-free electron–ion differential cross-sections (DCS) from the measured LIED momentum distribution. The required deconvolution procedure is based on the quantitative rescattering theory^{11,22,23}, previously validated for various atoms and molecules^{24–27}. It states that in the strong-field approximation, the detected momentum \mathbf{p} , shown in Fig. 1a for an unaligned N_2 distribution, is the vector sum of the momentum after recollision, \mathbf{p}_r , and the field's vector potential, \mathbf{A}_r , at the moment of rescattering, t_r . Electrons detected at various angles along the circumference defined at constant $|\mathbf{p}_r|$ (magenta dashed circle in Fig. 1a) thus constitute field-free DCS, equivalent to the DCS from CED. The DCS extracted according to this prescription from the experimentally determined momentum distribution for $|\mathbf{p}_r| = 2.71$ a.u. (atomic unit of linear momentum = 1.99×10^{-24} kg m s^{−1}) or 100-eV recollision energy is shown in Fig. 1b, and is found to agree well (within normalization constants) with the CED measurement²⁸ and our theoretical result. Consistent agreement is achieved for analysis at different laser wavelengths and intensities, and also for O_2 . (Supplementary Information provides a detailed description of the DCS extraction procedure and a presentation of the factors that influence the interpretation of the LIED momentum distribution, for example the role of short and long trajectories, multiple returns, inelastic scattering and molecular excitations.)

When bond lengths are being extracted from CED measurements, molecules are treated as sets of independent atoms with fixed interatomic distances; chemical properties, such as molecular bonding, are neglected. The validity of this so-called independent atom model¹² (IAM) approach depends on the dominance of core-penetrating collisions facilitated by multi-keV electron beams. Within IAM and restricting our discussion to homonuclear diatomic molecules, the DCS for electrons colliding with unaligned molecules can be expressed as $\sigma_M = 2\sigma_A + 2\sigma_A \sin(qR)/qR$. The first term on the right is the contribution to the DCS from the incoherent sum of two atoms and conveys no structural information. The second is a molecular interference term depending explicitly on the internuclear distance R and the elastic momentum transfer

¹Department of Physics, The Ohio State University, Columbus, Ohio 43210, USA. ²Department of Physics, Kansas State University, Manhattan, Kansas 66506, USA. ³Department of Chemistry, The Ohio State University, Columbus, Ohio 43210, USA.

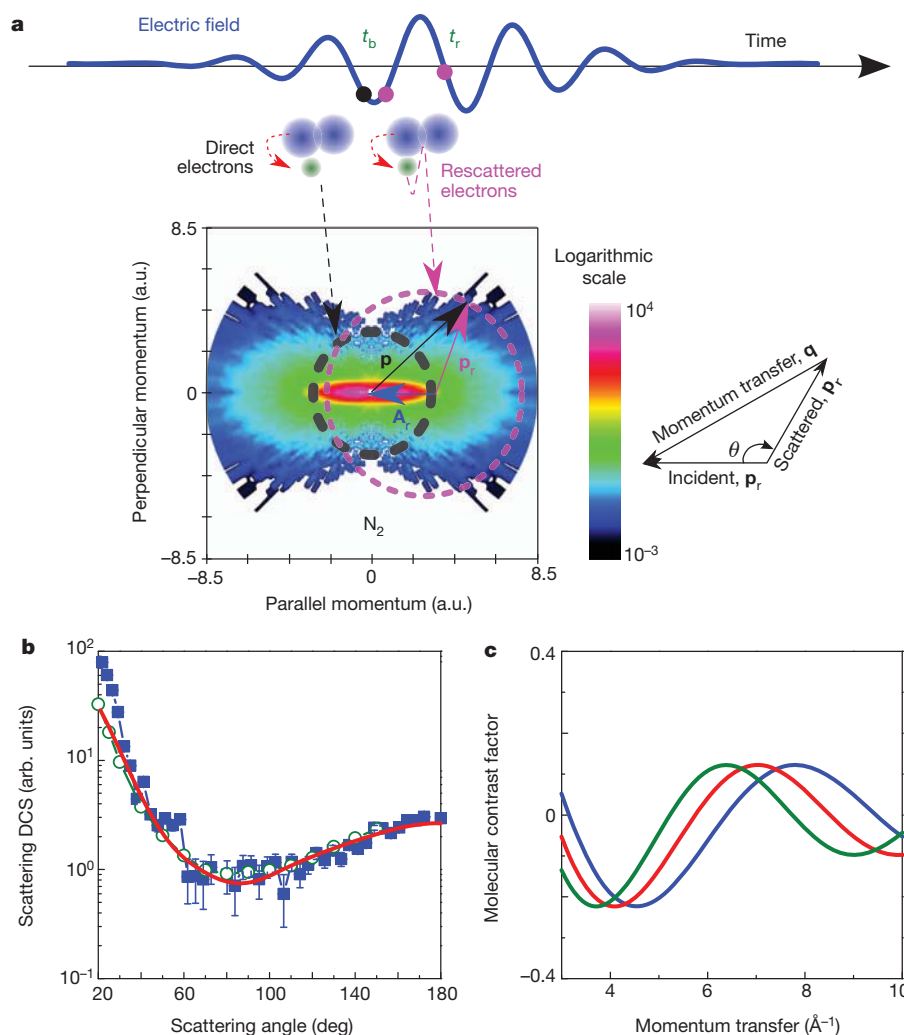


Figure 1 | The principle of laser-induced electron diffraction. **a**, Strong-field rescattering. An electron is released at time t_b in the electric field (solid blue line) of an intense mid-infrared laser. For certain initial times t_b , 'direct' electrons drift away from the core (solid blue), detected with momenta inside the dashed black circle. At other birth times, the electron returns roughly an optical cycle later back to the core (magenta trajectory). After elastic backscattering the electron gains additional momentum from the field, resulting in a larger detected momentum and separability from the direct ones (magenta dashed circle). A quasi-classical analysis allows one to deconstruct the detected momentum \mathbf{p} into the momentum during rescattering \mathbf{p}_r using the vector relationship $\mathbf{p} = \mathbf{p}_r - \mathbf{A}_r$ (atomic units), where $\mathbf{A}_r = (\mathbf{A})(t_r)$ is the field's vector potential at t_r . The inset defines the incident momentum \mathbf{p} , the scattered momentum \mathbf{p}_r , the rescattering angle θ and the momentum transfer \mathbf{q} . **b**, The

DCS (blue solid squares) extracted from the N_2 distribution in **a** (260 TW cm^{-2} , 50 fs, $2.0 \text{ }\mu\text{m}$ pulses) by sweeping the rescattering angle θ along the circumference of a circle of radius $|\mathbf{p}_r| = 2.71 \text{ a.u.}$ The error bars indicate the Poissonian uncertainty associated with the number of electrons detected at a given angle. The open green circles are data from CED measurements from ref. 28. The solid red line is calculated by using the independent atom model (IAM) for the CED data. At small scattering angles the LIED data are 'contaminated' by the direct electrons and are therefore not part of the analysis. **c**, Calculated MCFs plotted as a function of momentum transfer for three different N_2 bond distances: 0.99 \AA (blue curve), 1.10 \AA (red curve) and 1.21 \AA (green curve). The plot shows that as the internuclear separation R is changed, the fringes shift from $\sim 6 \text{ \AA}^{-1}$ to 8 \AA^{-1} .

$q = 2p_r \sin(\theta/2)$, where θ is the scattering angle shown in Fig. 1a. Good spatial resolutions require large momentum transfers producing small interference terms. In CED it is convenient to define a molecular contrast factor (MCF) as the ratio of the second term to the first term. For typical q values of about $5\text{--}10 \text{ \AA}^{-1}$, the MCF is ± 0.15 . With N_2 as an example, the method's spatial precision is emphasized by the extreme sensitivity of the MCF term to small changes in bond length, as shown in Fig. 1c by the clear shift in fringe maximum from 6 \AA^{-1} to 8 \AA^{-1} for $0.1\text{-}\text{\AA}$ changes in R . This is a well-known benefit in precision of any interferometric measurement, such as the Young double-slit experiment. Therefore, for large momentum transfers, bond lengths can be extracted with the use of LIED by using similar principles to CED (see Supplementary Information). In our experiment, sufficiently energetic collisions are generated by exploiting the λ^2 scaling of the rescattering process at long wavelengths^{15,16}. For instance, a $2.3\text{-}\mu\text{m}$ pulse results in an eightfold increase in rescattering energies over that

achieved with a traditional $0.8\text{-}\mu\text{m}$ Ti:sapphire laser, yielding rescattering energies in excess of 200 eV. In addition, our LIED analysis is restricted to large-angle scattering, thus producing large momentum transfers, comparable in value to those of CED.

The dynamics of the rescattering process implies that, at constant $|\mathbf{p}_r|$, the time Δt between birth and rescattering of the electron wave packet is proportional to the laser optical cycle and hence the wavelength. Consequently, after ionization at t_b , the bond length begins adjusting to the new electronic configuration and thus at t_r the returning wave packet captures the diffraction image corresponding to a non-equilibrium internuclear distance. Thus, a series of images captured at different wavelengths equates to mapping the nuclear motion. If the electron removal is rapid, the Frank-Condon principle can be applied to evaluate the subsequent motion of the vibrational wave packet. We emphasize that Frank-Condon estimates are only approximate, because the impulsive nature of tunnel ionization is speculative. In this

study, the three mid-infrared wavelengths were, in principle, probing the internuclear separation 4–6 fs after tunnelling ionization.

The baseline for our LIED investigation was ionization of the highest occupied molecular orbital (HOMO) of N_2 , the bonding σ_a orbital. Removal of the σ_a electron results in a small change in the equilibrium N–N distance from 1.10 Å (neutral) to 1.12 Å (ion)²⁹. The experimental DCS (for example Fig. 1b) is matched to an IAM calculation by fitting the normalization constant and the internuclear separation so as to minimize the variance between experiment and theory (see Supplementary Information). Figure 2a shows the best-fit MCF (red line) retrieved from the 2- μ m experiment (symbols), obtained with a N–N distance of 1.14 Å. Comparing CED data at the same collision energy

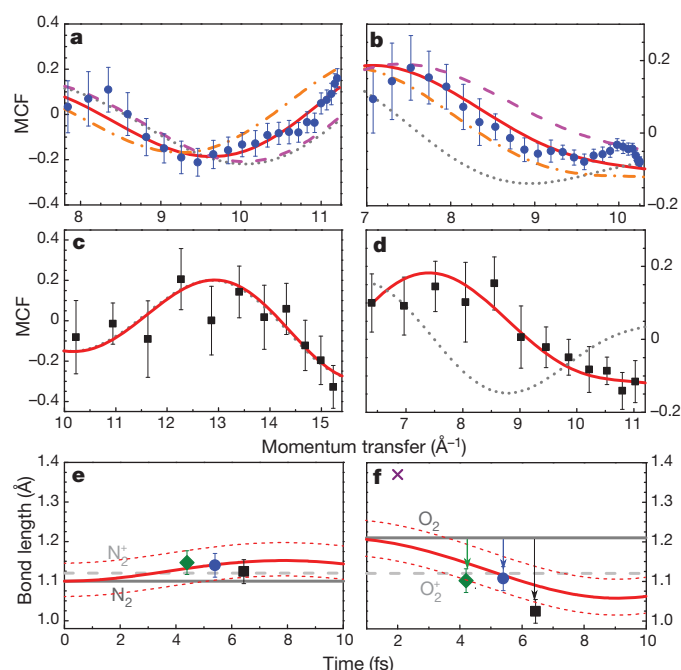


Figure 2 | LIED for unaligned N_2 and O_2 molecules. **a–d**, MCF extracted from the experimental data (scattered symbols) in comparison with theoretical predictions. The best-fit bond lengths are obtained by fitting the DCS extracted directly from the measurement (see Supplementary Information). The solid red line is the MCF calculated using the best-fit bond length. The dotted grey line is the MCF calculated using the equilibrium N_2 bond length. The dashed magenta and dash-dotted orange lines in **a** and **b** are the calculated MCFs using bond lengths that deviate by -5 and $+5$ pm, respectively, around the best fit. The error bars obey a Poisson statistical distribution. **a**, N_2 data taken with 260 TW cm^{-2} , $2.0 \mu\text{m}$ pulses ($p_r = 2.97 \text{ a.u.}$). **c**, N_2 data taken with 290 TW cm^{-2} , $2.3 \mu\text{m}$ pulses ($p_r = 4.11 \text{ a.u.}$). **b**, O_2 data taken with 133 TW cm^{-2} , $2.0 \mu\text{m}$ pulses ($p_r = 2.91 \text{ a.u.}$). **d**, O_2 data taken with 150 TW cm^{-2} , $2.3 \mu\text{m}$ pulses ($p_r = 2.97 \text{ a.u.}$). **e**, **f**, Illustration of bond change for N_2 (**e**) and O_2 (**f**) after ionization. The bond lengths extracted from the LIED measurements at each wavelength (squares, $2.3 \mu\text{m}$; circles, $2 \mu\text{m}$; diamonds, $1.7 \mu\text{m}$) are shown. The associated error bars represent the IAM standard deviation resulting from experimental uncertainties. The time and wavelength correspondence is based on a classical analysis (see Supplementary Information). The red curves depict the evolution of the nuclear wave packet's centre (solid) and its associated full width at half-maximum (dotted), computed in the Frank–Condon approximation. The equilibrium bond lengths are indicated by the solid (neutral species) and dashed grey (ion) lines. For O_2 , the vertical arrows indicate that the measured bond lengths for all three wavelengths are consistently shorter ($\sim 0.1 \text{ Å}$) than the equilibrium length for the neutral species and are statistically meaningful. The classical behaviour suggests that this is a consequence of the 4–6-fs bond evolution after ionization. The bond length reported in ref. 10 (purple cross) was extracted from a $0.8 \mu\text{m}$ laser experiment on aligned O_2 molecules. The short wavelength produced smaller return energies ($< 50 \text{ eV}$), thus complicating the interpretation and analysis of the momentum distribution, ultimately producing larger uncertainty in bond length.

(120 eV), we estimate a 0.05-Å error on the retrieved bond length. For comparison, we also plot in Fig. 2a the theoretical MCF calculated for N–N distances deviating from the fitted value by $\pm 5 \text{ pm}$ (dashed lines) and equal to that of neutral N_2 (dotted line). Figure 2c shows the analysis performed for a collision momentum of 4.11 a.u. (230 eV energy) using $2.3\text{-}\mu\text{m}$ pulses at 290 TW cm^{-2} . The retrieved N–N distance is 1.12 Å . In addition, LIED measurements performed at $1.7 \mu\text{m}$ result in a retrieved bond length of 1.15 Å . These three retrieved bond lengths are plotted in Fig. 2e as a function of time, using the classical correspondence with wavelength discussed above. The three retrieved bond lengths agree, within a 0.05-Å uncertainty, with the accepted N_2 equilibrium value. However, the experiment cannot resolve the small N–N bond motion ($\sim 0.02 \text{ Å}$) during the 4–6-fs time interval after tunnel ionization.

In contrast to N_2 , the O_2 HOMO is an anti-bonding π_a orbital, and ionization leads to a large change in the O–O equilibrium distance from 1.21 Å (neutral) to 1.10 Å (ion)²⁹. By following the same procedure as outlined above, the extracted O–O distances for 1.7 , 2.0 and $2.3\text{-}\mu\text{m}$ LIED measurements are 1.10 , 1.11 and 1.02 Å , respectively. Figure 2b, d shows the MCF analysis at the last two wavelengths, again by comparing experimentally determined MCF values against the theoretical best fit, and against theoretical fits obtained for a bond length that differs from the best fit by $\pm 5 \text{ pm}$ or is equal to that of neutral O_2 . In contrast to the N_2 experiments, the MCF curve calculated by using the neutral equilibrium distance does not fit the experiment at either wavelength. The O–O distances derived from the best fits deviate by 2–4 standard deviations from the equilibrium value, providing a high degree of confidence ($> 99.9\%$) that the bond has shortened in the 4–6-fs interval after ionization. The Frank–Condon curves in Fig. 2f show the free time evolution of the centre of the nuclear wave packet (O–O distance, solid red line), whereas dotted red lines indicate its width (full width at half-maximum). The three experimental data points in Fig. 2f show a statistically significant decrease (0.1 Å) in the O–O bond length from its initial, neutral equilibrium distance of 1.21 Å , as indicated by the vertical arrows. However, the spatial resolution of the experiment is insufficient to track the much smaller change in bond length ($\sim 0.05 \text{ Å}$ resolution) occurring within the 4–6-fs time window spanned by the three measurement points. Regarding temporal resolution, we estimate an uncertainty of 2–3 fs in assigning a return time for each wavelength (see Supplementary Information).

Our findings illustrate that LIED can image sub-ångström structural changes in gas-phase molecules with femtosecond time resolution. We expect that the use of few-cycle pulses, bichromatic fields, additional wavelengths and improved counting rates will enhance the temporal resolution, and that the method can be used in pump-probe setups to interrogate molecules undergoing conformational transformations. A next step would be to apply LIED to more complex molecules, to aligned or oriented molecules, and to vibrationally or electronically excited molecules. Compared with a CED experiment, LIED's ability to control the returning electron wave packet coherently by means of bichromatic laser fields and/or ellipticity could open new paths in electron diffraction.

METHODS SUMMARY

Details of the laser system used in the experiment have been described previously¹⁶. In brief, an optical parametric amplifier (Topas-HE; LightConversion) pumped by a near-infrared Ti:sapphire front-end subsystem ($0.8 \mu\text{m}$, 4.4 mJ , 50 fs , 1 kHz) generates continuously tunable ultra-short (0.5 mJ , 50 fs) linearly polarized pulses from 1.2 to $2.3 \mu\text{m}$. Angle-resolved (collection angle 1.6°) photoelectron energy distributions were recorded by focusing the laser in a $\sim 22\text{-cm}$ field-free time-of-flight electron spectrometer. At each wavelength, the laser-driven diffraction image was obtained by rotating the laser polarization with respect to the spectrometer axis in steps of 2° , using zero-order half-wave plates. The sequence of collection angles was randomized to minimize systematic errors. For each angle, data were collected for about 10^6 laser shots, thus ensuring the necessary dynamic range while the count rate was kept below three hits per shot to minimize space-charge effects.

The energy calibration of the spectrometer was performed by recording the well-known above-threshold ionization (ATI) spectra of various noble gases with long (230-ps) pulses. The relative energy resolution $\Delta E/E$ of the apparatus was determined to be better than 1.4%, inferred from the long-lived Rydberg series present in the ATI spectrum of argon taken with 50-fs, 0.8- μ m pulses. Finally, for each wavelength the peak intensities were estimated by analysing the $2U_p$ and $10U_p$ classical cutoffs present in the photoelectron energy distribution (see Supplementary Information for details). The cutoff values agree very well with numerical results based on single-active electron quantum calculations¹⁶, and we estimate the accuracy of our technique to be within an error of 15%. These uncertainties were found to lead to 1–2-pm changes in the extracted bond lengths for all data sets.

Received 8 June; accepted 15 December 2011.

- Seibert, M. M. *et al.* Single mimivirus particles intercepted and imaged with an X-ray laser. *Nature* **470**, 78–81 (2011).
- Chapman, H. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73–77 (2011).
- Zewail, A. H. & Thomas, J. M. *4D Electron Microscopy: Imaging in Space and Time* (Imperial College Press, 2010).
- Eichberger, M. *et al.* Snapshots of cooperative atomic motions in the optical suppression of charge density waves. *Nature* **468**, 799–802 (2010).
- Sciaini, G. & Miller, D. Femtosecond electron diffraction: heralding the era of atomically resolved dynamics. *Rep. Prog. Phys.* **74**, 096101 (2011).
- Elsaesser, T. & Woerner, M. Photoinduced structural dynamics of polar solids studied by femtosecond X-ray diffraction. *Acta Crystallogr. A* **66**, 168–178 (2010).
- Niikura, H. *et al.* Sub-laser-cycle electron pulses for probing molecular dynamics. *Nature* **417**, 917–922 (2002).
- Itatani, J. *et al.* Tomographic imaging of molecular orbitals. *Nature* **432**, 867–871 (2004).
- Li, W. *et al.* Time-resolved dynamics in N_2O_4 probed using high harmonic generation. *Science* **322**, 1207–1211 (2008).
- Meckel, M. *et al.* Laser-induced electron tunneling and diffraction. *Science* **320**, 1478–1482 (2008).
- Lin, C. D., Le, A. T., Chen, Z., Morishita, T. & Lucchese, R. Strong-field rescattering physics—self-imaging of a molecule by its own electrons. *J. Phys. B* **43**, 122001 (2010).
- Hargittai, I. & Hargittai, M. (eds) *Stereochemical Applications of Gas-Phase Electron Diffraction* (VCH, 1998).
- Reckenthaeler, P. *et al.* Time-resolved electron diffraction from selectively aligned molecules. *Phys. Rev. Lett.* **102**, 213001 (2009).
- Zewail, A. H. Femtochemistry: atomic-scale dynamics of the chemical bond using ultrafast lasers. *Angew. Chem. Int. Ed.* **39**, 2586–2631 (2000).
- Tate, J. *et al.* Scaling of wave-packet dynamics in an intense midinfrared field. *Phys. Rev. Lett.* **98**, 013901 (2007).
- Colosimo, P. *et al.* Scaling strong-field interactions towards the classical limit. *Nature Phys.* **4**, 386–389 (2008).
- Xu, J., Chen, Z., Le, A. T. & Lin, C. D. Self-imaging of molecules from diffraction spectra by laser-induced rescattering electrons. *Phys. Rev. A* **82**, 033403 (2010).
- Schafer, K. J., Yang, B., DiMauro, L. F. & Kulander, K. C. Above threshold ionization beyond the high harmonic cutoff. *Phys. Rev. Lett.* **70**, 1599–1602 (1993).
- Corkum, P. B. Plasma perspective on strong field multiphoton ionization. *Phys. Rev. Lett.* **71**, 1994–1997 (1993).
- Baker, S. *et al.* Probing proton dynamics in molecules on an attosecond time scale. *Science* **312**, 424–427 (2006).
- Torres, R. *et al.* Revealing molecular structure and dynamics through high-order harmonic generation driven by mid-IR fields. *Phys. Rev. A* **81**, 051802(R), <http://dx.doi.org/10.1103/PhysRevA.81.051802> (2010).
- Morishita, T., Le, A. T., Chen, Z. & Lin, C. D. Accurate retrieval of structural information from laser-induced photoelectron and high-order harmonic spectra by few-cycle laser pulses. *Phys. Rev. Lett.* **100**, 013903 (2008).
- Chen, Z., Le, A. T., Morishita, T. & Lin, C. D. Quantitative rescattering theory for laser-induced high-energy plateau photoelectron spectra. *Phys. Rev. A* **79**, 033409 (2009).
- Ray, D. *et al.* Large-angle electron diffraction structure in laser-induced rescattering from rare gases. *Phys. Rev. Lett.* **100**, 143002 (2008).
- Okunishi, M. *et al.* Experimental retrieval of target structure information from laser-induced rescattered photoelectron momentum distributions. *Phys. Rev. Lett.* **100**, 143001 (2008).
- Okunishi, M., Niikura, H., Lucchese, R. R., Morishita, T. & Ueda, K. Extracting electron-ion differential scattering cross sections for partially aligned molecules by laser-induced rescattering photoelectron spectroscopy. *Phys. Rev. Lett.* **106**, 063001 (2011).
- Cornaggia, C. Electron-ion elastic scattering in molecules probed by laser-induced ionization. *J. Phys. B* **42**, 161002 (2009).
- Dubois, R. D. & Rudd, M. E. Differential cross sections for elastic scattering of electrons from argon, neon, nitrogen and carbon monoxide. *J. Phys. B* **9**, 2657–2667 (1976).
- Dekock, R. L. & Gray, H. B. *Chemical Structure and Bonding* 229 (University Science Books, 1989).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The work at The Ohio State University and Kansas State University was performed under DOE/BES contracts DE-FG02-06ER15833 and DE-FG02-06ER15832, respectively. L.F.D. acknowledges support from the Hagenlocker chair.

Author Contributions C.I.B. and J.X. designed the experiment and performed the data analysis. C.I.B., A.D.D., E.S. and K.Z. performed the experiment. J.X. and C.D.L. performed the theoretical analysis. C.I.B., J.X., A.D.D., P.A., T.A.M., L.F.D. and C.D.L. interpreted the results and prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.I.B. (cbaga@mps.ohio-state.edu).

Stability criteria for complex ecosystems

Stefano Allesina^{1,2} & Si Tang¹

Forty years ago, May proved^{1,2} that sufficiently large or complex ecological networks have a probability of persisting that is close to zero, contrary to previous expectations^{3–5}. May analysed large networks in which species interact at random^{1,2,6}. However, in natural systems pairs of species have well-defined interactions (for example predator–prey, mutualistic or competitive). Here we extend May's results to these relationships and find remarkable differences between predator–prey interactions, which are stabilizing, and mutualistic and competitive interactions, which are destabilizing. We provide analytic stability criteria for all cases. We use the criteria to prove that, counterintuitively, the probability of stability for predator–prey networks decreases when a realistic food web structure is imposed^{7,8} or if there is a large preponderance of weak interactions^{9,10}. Similarly, stability is negatively affected by nestedness^{11–14} in bipartite mutualistic networks. These results are found by separating the contribution of network structure and interaction strengths to stability. Stable predator–prey networks can be arbitrarily large and complex, provided that predator–prey pairs are tightly coupled. The stability criteria are widely applicable, because they hold for any system of differential equations.

May's theorem deals with community matrices^{1,2,6} M , of size $S \times S$, where S is the number of species. M_{ij} describes the effect that species j has on i around a feasible equilibrium point (that is, species have positive densities) of an unspecified dynamical system describing the species' densities through time.

In May's work^{1,2}, the diagonal coefficients are -1 , and the off-diagonal coefficients are drawn from a distribution with mean 0 and variance σ^2 with probability C and are 0 otherwise. For these matrices, the probability of stability is close to 0 whenever the 'complexity' $\sigma\sqrt{SC} > 1$. Local stability measures the tendency of the system to return to equilibrium after perturbations. In unstable systems, even infinitesimal perturbations cause the system to move away from equilibrium, potentially leading to the loss of species. Thus, it should be extremely improbable to observe rich (large S) or highly connected (large C) persistent ecosystems^{1,2}. Mathematically, an equilibrium point is stable if all the eigenvalues of the community matrix have negative real parts^{1,2,6}.

Local stability can only describe the behaviour of the system around an equilibrium point, whereas natural systems are believed to operate far from a steady state^{5,15}. However, methods based on local stability are well suited to the study of large systems^{1,16,17}, whose empirical parameterization would be unfeasible. Moreover, the methods are general, so that they can be applied to any system of differential equations.

May's matrices have random structure: each pair of species interacts with the same probability. However, this randomness translates, for large S , into fixed interaction frequencies, so that these matrices follow a precise mixture of interaction types. For example, in May's matrices predator–prey interactions are twice as frequent as mutualistic ones (Supplementary Table 1). Here we extend May's work to different types of interaction, starting from the random case.

Suppose that two species j and i interact with probability C , and that the interaction strength is drawn from a distribution: M_{ij} takes the value of a random variable X with mean $\mathbb{E}(X)=0$ and variance

$\text{Var}(X) = \sigma^2$. The diagonal elements of the community matrix, representing self-regulation, are set to $-d$. For large systems, the eigenvalues are contained in a circle¹⁸ in the complex plane (Fig. 1 and Supplementary Information). The circle is centred at $(-d, 0)$ and the radius is $\sigma\sqrt{SC}$. In stable systems, the whole circle is contained in the left half-plane (that is, all eigenvalues have negative real parts). Thus, the system is stable when the radius is smaller than d : $\sqrt{SC} < \theta = d/\sigma$.

In predator–prey networks, interactions come in pairs with opposite signs: whenever $M_{ij} > 0$, then $M_{ji} < 0$. With probability C , we sample one interaction strength from the distribution of $|X|$ and the other from $-|X|$, whereas with probability $(1 - C)$ both are zero. The eigenvalues of large predator–prey matrices are contained in a vertically stretched ellipse¹⁹, centred at $(-d, 0)$, with horizontal radius $\sigma\sqrt{SC}(1 - \mathbb{E}^2(|X|)/\sigma^2)$ and thus the stability criterion is $\sqrt{SC} < \theta/(1 - \mathbb{E}^2(|X|)/\sigma^2)$ (Fig. 1 and Supplementary Information).

When we constrain M_{ij} and M_{ji} to have the same sign, and thus impose a mixture of competition and mutualism with equal probability, the eigenvalues are enclosed in a horizontally stretched ellipse¹⁹ and the criterion becomes $\sqrt{SC} < \theta/(1 + \mathbb{E}^2(|X|)/\sigma^2)$ (Fig. 1 and Supplementary Information).

Take $C = 0.1$, $X \sim N(0, 1/4)$ (that is, X follows a normal distribution with mean 0 and variance $1/4$), and $d = 1$. The criterion becomes $\sqrt{SC} < 2$ for random matrices, and is violated whenever $S \geq 41$. For predator–prey we find $\sqrt{SC} < 2\pi/(\pi - 2)$ (violated for $S \geq 303$) and for the mixture of competition and mutualism $\sqrt{SC} < 2\pi/(\pi + 2)$ (violated for $S \geq 15$). Since $\mathbb{E}(|X|)/\sigma > 0$ for any distribution of X , the stability criteria form a strict hierarchy in which the mixture matrices are the least likely to be stable, the random matrices are intermediate, and the predator–prey matrices are the most likely to be stable (Fig. 2 and Table 1). Considerations based on qualitative stability² and numerical simulations¹⁶ are consistent with this hierarchy.

In the three cases above, the mean interaction strength is zero, and the coefficients come from the same distribution. In fact we can shuffle the interaction strengths, thereby transforming a network of one type into another: the difference in stability is driven exclusively by the arrangement of the coefficients in pairs with random, opposite and same signs, respectively. This feature allows us to further derive the stability criteria for all intermediate cases by using linear combinations of the three cases above (Supplementary Information).

Two ecologically important cases, however, cannot produce a mean interaction strength of zero. In mutualistic networks all interactions are positive, whereas in competitive networks they are negative. In these cases, for large systems, all the eigenvalues except one (equal to the row sum) are contained in an ellipse (Fig. 3 and Supplementary Figs 1 and 2). In mutualistic networks in which all interaction pairs are positive and drawn from the distribution of $|X|$ independently with probability C , the stability criterion becomes $(S - 1)C\mathbb{E}(|X|)/\sigma < \theta$ (that is, row sum < 0 ; Supplementary Information). For competitive matrices, in which interaction pairs are drawn from the distribution of $-|X|$ with probability C , the criterion is

$$\sqrt{SC}(1 + (1 - 2C)\mathbb{E}^2(|X|)/\sigma^2) / \sqrt{1 - C\mathbb{E}^2(|X|)/\sigma^2 + C\mathbb{E}(|X|)/\sigma} < \theta$$

¹Department of Ecology and Evolution, University of Chicago, 1101 East 57th Street, Chicago, Illinois 60637, USA. ²Computation Institute, University of Chicago, 5735 South Ellis Avenue, Chicago, Illinois 60637, USA.

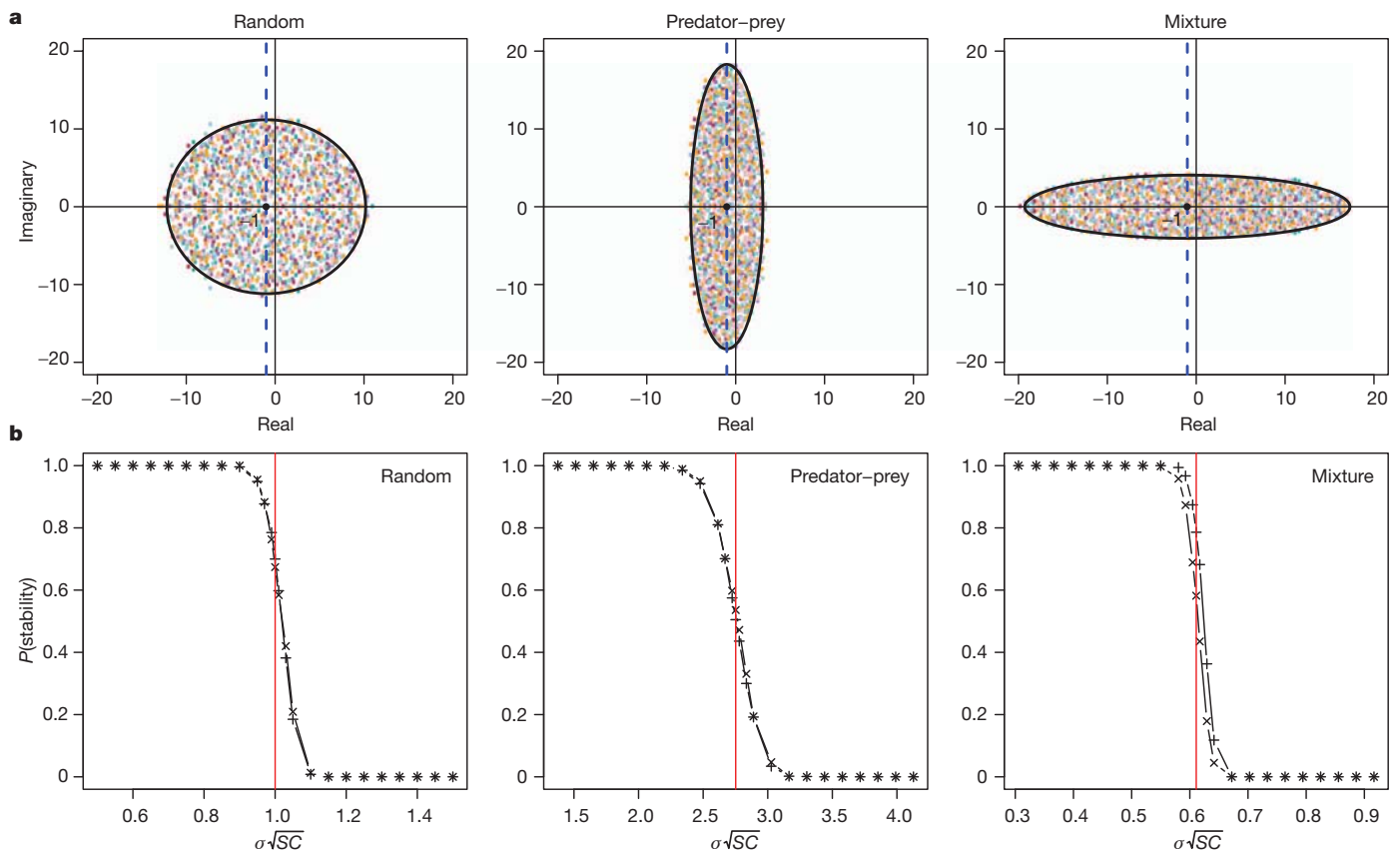


Figure 1 | Distributions of the eigenvalues and corresponding stability profiles. **a**, For $X \sim N(0, \sigma^2)$, $S = 250$, $C = 0.25$ and $\sigma = 1$, we plot the eigenvalues of 10 matrices (colours) with $-d = -1$ on the diagonal and off-diagonal elements, following the random, predator-prey or mixture prescriptions. The black ellipses are derived analytically in the text. **b**, Numerical simulations for the corresponding stability profiles. For the random case, starting from $S = 250$, $C = 0.5$, $\sigma = 0.1$ and $d = 1$, we

systematically varied C (crosses) or σ (plus signs) to obtain $\sigma\sqrt{SC}$ spanning $[0.5, \dots, 1.0, \dots, 1.5]$ of the critical value for stability (indicated in red, 1 in the case of random matrices). The profiles were obtained by computing the probability of stability out of 1,000 matrices. The predator-prey case is as the random but with $\sigma = 0.5$ and critical value $\pi/(\pi - 2)$. The mixture case is as the random but with critical value $\pi/(\pi + 2)$. In all cases, the phase transition between stability and instability is accurately predicted by our derivation.

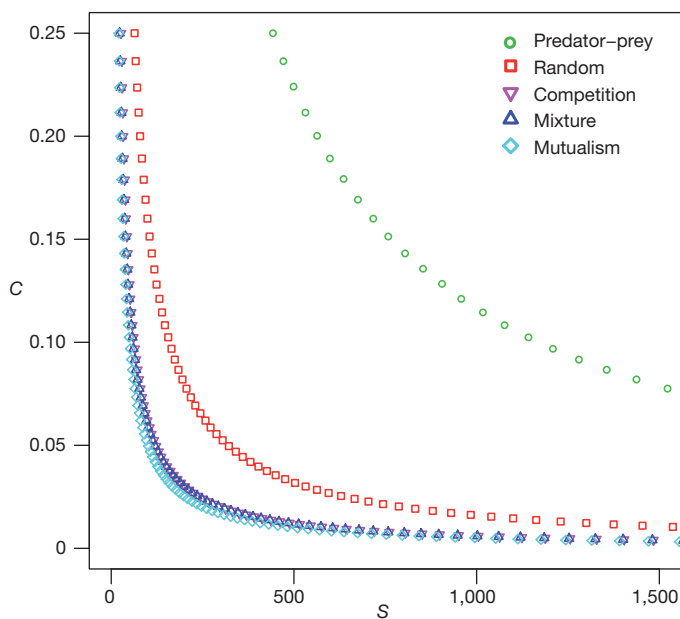


Figure 2 | Stability criteria for different types of interaction. We fixed $\theta = d/\sigma = 4$, and for a given connectance C we solved for the largest integer S that satisfies the stability criterion for each type of interactions. Combinations of S and C below each curve lead to stable matrices with a probability close to 1. The interaction types form a strict hierarchy from mutualism (most unlikely to be stable) to predator-prey (most likely to be stable).

(Supplementary Information). In both cases, stability decreases rapidly with higher complexity, and mutualistic matrices are less likely to be stable than their competitive counterpart (Fig. 2 and Table 1).

Having derived the stability criteria, we want to assess the effect of imposing realistic food web structure within the predator-prey case. It is believed that realistic food web structures should improve stability^{7,8,17}. In community matrices of food webs, producers have positive columns and negative rows, with the opposite for top predators. To test whether these variations affect stability, we plotted the eigenvalues for predator-prey webs in which interactions are arranged, following the cascade²⁰ and niche²¹ models. Imposing realistic structures results in eigenvalues with larger real parts than the corresponding unstructured predator-prey case (Supplementary Information and Supplementary Fig. 3). Thus, the cascade and niche models produce networks that are less likely to be stable than their unstructured predator-prey counterpart, with the niche model having a larger discrepancy: imposing realistic food web structure hampers stability.

Similarly, we measured the effect of realistic structures on mutualistic networks. Several published mutualistic networks are bipartite^{11–14}: there are two types of node (for example plants and pollinators), and interactions occur exclusively between different types. In addition, bipartite mutualistic networks tend to be nested¹¹: the interactions of the specialists form a subset of those of the generalists. Nestedness is believed to beget stability^{12–14}. We plotted the eigenvalues for these two types of structure and compared the results with those obtained for the unstructured mutualistic case (Fig. 3, Supplementary Information and Supplementary Fig. 4). As stated above, stability in mutualistic networks is determined by the row sum. The bipartite case yields row sums that,

Table 1 | Stability criteria for different types of interaction and network structure

Interaction	Stability criterion	$S_{\max}(C, \theta)$		
		(0.1, 2.0)	(0.1, 4.0)	(0.2, 4.0)
Nested mutualism	$(S-1)C\sqrt{\frac{2}{\pi}} < \theta$	9	28	18
Mutualism		16 (15)	41 (51)	22 (20)
Bipartite mutualism	$\sqrt{SC} < \frac{\theta\pi}{\pi+2}$	17	41	23
Mixture		17 (14)	58 (59)	33 (29)
Competition	$\sqrt{SC}\left(1 + \frac{2-2C}{\pi-2C}\right)\sqrt{\frac{\pi-2C}{\pi}} + C\sqrt{\frac{2}{\pi}} < \theta$	17 (15)	62 (63)	38 (33)
Random		50 (40)	168 (160)	88 (80)
Niche predator–prey	$\sqrt{SC} < \frac{\theta\pi}{\pi-2}$	149	461	245
Cascade predator–prey		298	1,134	535
Predator–prey		314 (302)	1,201 (1,211)	603 (605)

In all cases, the criterion is derived for large $S \times S$ matrices with $X \sim N(0, \sigma^2)$ (and thus $\mathbb{E}(|X|) = \sigma\sqrt{2/\pi}$), connectance C and $\theta = d/\sigma$. Numerical simulations report, for a given combination of C and θ , the largest S (S_{\max}) yielding a probability of stability ≥ 0.5 (computed using 1,000 matrices). In parenthesis are the analytical predictions.

for large S , are equal to the unstructured case. Accordingly, we did not find a discrepancy in stability for the bipartite case. However, in nested structures some rows and columns have sums that are larger than average (generalist plants and animals). Consequently, nested matrices are inherently less likely to be stable than unstructured ones. These findings are confirmed by numerical simulations. Using the same method, we found that asymmetric coupling of interaction strengths (where each large M_{ij} is coupled with a small M_{ji}), contrary to current expectations²², does not influence stability in mutualistic networks (Supplementary Information and Supplementary Fig. 5).

We have considered how the arrangement of the interactions affects stability, and have found several counterintuitive results. These results can be accounted for by the fact that we provide a very conservative test for the effects of structure on stability (Supplementary Information). We now assess the role of the magnitude of interaction strengths. In fact, our findings extend to any distribution of coefficient strengths (Supplementary Information).

Typically, ecologists have regarded σ as the ‘average interaction strength’^{11,2}. However, σ does not provide information on weak interactions^{9,10,17}: we can have the same σ for two distributions with distinct shapes, and thus different proportions of weak and strong interactions (Supplementary Information). We analyse how the shape of the distribution affects stability for fixed S , C , d and σ . If the distribution contains many weak interactions, the expected magnitude $\mathbb{E}(|X|) \approx 0$. In contrast, if weak interactions are rare, $\mathbb{E}(|X|) \approx \sigma$. In the predator–prey systems, lowering $\mathbb{E}(|X|)$ decreases $\theta/(1 - \mathbb{E}^2(|X|)/\sigma^2)$ and thus

hampers stability. We conclude that weak interactions, contrary to current beliefs^{9,10,17}, can destabilize predator–prey systems. Weakening the interactions shifts $\mathbb{E}(|X|)$ closer to zero and therefore makes predator–prey systems closer to their random counterpart. With the same argument, weak interactions can stabilize the mixture of competition and mutualism case and have no effect on random networks. Variability in interaction strengths was previously found to be detrimental for stability in large food webs²³ and competitive networks¹⁷.

For example, consider a uniform distribution $X \sim U[-\sigma\sqrt{3}, \sigma\sqrt{3}]$ and contrast it with the normal case $X \sim N(0, \sigma^2)$. Both parameterizations lead to $\mathbb{E}(X) = 0$ and $\text{Var}(X) = \sigma^2$. In the uniform case, $\mathbb{E}(|X|) = \sigma\sqrt{3}/2 \approx 0.866\sigma$, whereas in the normal case $\mathbb{E}(|X|) = \sigma\sqrt{2/\pi} \approx 0.798\sigma$. This means that the uniform distribution, on average, leads to stronger interactions than the corresponding normal case. In turn, this has a large effect on stability: the criterion for the predator–prey case becomes $\sqrt{SC} < 4\theta$ for the uniform distribution, whereas it is $\sqrt{SC} < \pi/(\pi-2)\theta \approx 2.75\theta$ for the normal case. The random case is unaffected by the choice of the distribution ($\sqrt{SC} < \theta$), whereas in the mixture of competition and mutualism we have $\sqrt{SC} < 4\theta/7 \approx 0.571\theta$ for the uniform distribution and $\sqrt{SC} < \pi\theta/(\pi+2) \approx 0.61\theta$ for the normal case. These considerations extend to any choice of distribution for the interaction strengths (Supplementary Information and Supplementary Figs 6 and 7): weak interactions, all other things being equal, are destabilizing for food webs, stabilizing for mutualistic and competitive networks (and their mixture), and have no effect on random networks.

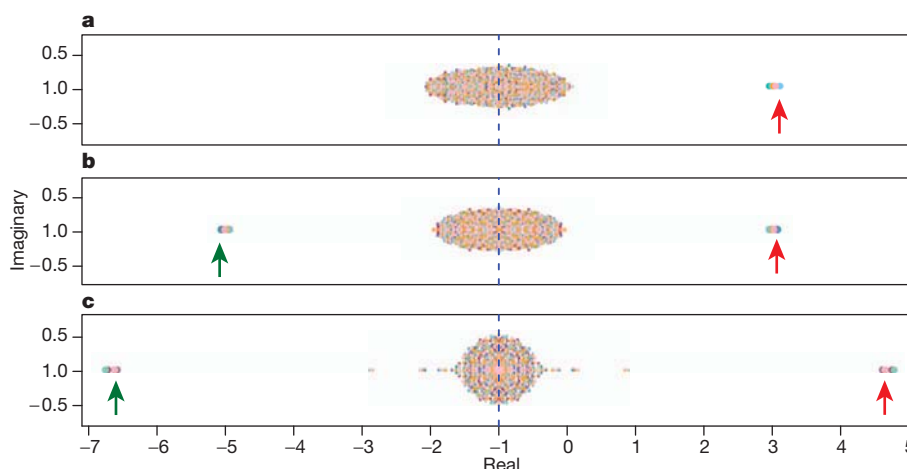


Figure 3 | Distribution of the eigenvalues for the three types of mutualism. **a**, Unstructured mutualism. **b**, Bipartite mutualism. **c**, Nested and bipartite mutualism. In all cases, $S = 250$, $\sigma = 0.1$, $C = 0.2$ and $d = 1$. Note that the bipartite case does produce extreme negative real eigenvalues (green arrow) coupled with positive ones, but the row sum (and thus the rightmost eigenvalue,

red arrow) is equal to that of the unstructured mutualistic case. The nested matrices, in which generalist species yield (on average) larger row and column sums, have larger rightmost eigenvalues. Thus, highly nested matrices are less likely than the other two cases to be stable.

We have derived stability criteria for unstructured networks in which species interact at random, in predator–prey, mutualistic, and competitive pairs. These results hold for arbitrary diagonal values and arbitrary distribution of interaction strengths (Supplementary Information). Our analysis shows that, all other things being equal, weak interactions can be either stabilizing or destabilizing depending on the type of interactions between species. In predator–prey systems, realistic structure and weak interactions are detrimental for stability. However, in natural food webs, which seem to persist in time, weak interactions are preponderant²⁴. The persistence of these networks might be explained by the interplay between their structure and weak interactions, even though each would be destabilizing if taken in isolation. For example, as suggested previously², generalist predators could have weak interactions with their numerous prey, reducing the effect of the realistic structure and driving the system closer to the unstructured case.

Predator–prey systems differ markedly from the other cases studied here. Suppose that a network is unstable. The system can be stabilized either by lowering C , S or σ (decreasing its complexity), or by increasing the self-regulation d . This is in line with May's argument: large and highly interconnected systems are difficult to stabilize. For random networks, reducing complexity is the only way to stabilize the system. However, in the other cases, networks can be stabilized by altering the distribution of interaction strengths; by modifying the parameters of the system we can typically change the distribution of the off-diagonal elements without altering the diagonal ones (Supplementary Information). For competition, mutualism and their mixture, stability is achievable by decreasing the average interaction strength $\mathbb{E}(|X|)$, which is akin to lowering complexity. On the contrary, predator–prey networks can be stabilized by increasing the strength of interaction $\mathbb{E}(|X|)$, and thus the coupling between predators and prey. Predator–prey systems are therefore the only ones that can potentially elude May's conclusions^{1,2} and support an arbitrarily large, complex and stable ecological network.

Our results show that the ubiquity of consumer–resource relationships in nature could be due to their intrinsic dynamical properties. These findings are not limited to ecological networks, but instead hold for any system of differential equations resting at an equilibrium point.

Received 18 May 2011; accepted 6 January 2012.

Published online 19 February 2012.

1. May, R. M. Will a large complex system be stable? *Nature* **238**, 413–414 (1972).
2. May, R. M. *Stability and Complexity in Model Ecosystems* (Princeton Univ. Press, 2001).

3. MacArthur, R. Fluctuations of animal populations and a measure of community stability. *Ecology* **36**, 533–536 (1955).
4. Elton, C. S. *Animal Ecology* (Univ. of Chicago Press, 2001).
5. McCann, K. S. The diversity–stability debate. *Nature* **405**, 228–233 (2000).
6. Levins, R. *Evolution in Changing Environments: Some Theoretical Explorations* (Princeton Univ. Press, 1968).
7. McNaughton, S. J. Stability and diversity of ecological communities. *Nature* **274**, 251–253 (1978).
8. Yodzis, P. The stability of real ecosystems. *Nature* **289**, 674–676 (1981).
9. McCann, K. S., Hastings, A. & Huxel, G. R. Weak trophic interactions and the balance of nature. *Nature* **395**, 794–798 (1998).
10. Emmerson, M. & Yearsley, J. M. Weak interactions, omnivory and emergent food-web properties. *Proc. R. Soc. Lond. B* **271**, 397–405 (2004).
11. Bascompte, J., Jordano, P., Melián, C. J. & Olesen, J. M. The nested assembly of plant–animal mutualistic networks. *Proc. Natl Acad. Sci. USA* **100**, 9383–9387 (2003).
12. Okuyama, T. & Holland, J. N. Network structural properties mediate the stability of mutualistic communities. *Ecol. Lett.* **11**, 208–216 (2008).
13. Bastolla, U. *et al.* The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature* **458**, 1018–1020 (2009).
14. Thébault, E. & Fontaine, C. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science* **329**, 853–856 (2010).
15. DeAngelis, D. L. & Waterhouse, J. C. Equilibrium and nonequilibrium concepts in ecological models. *Ecol. Monogr.* **57**, 1–21 (1987).
16. Allesina, S. & Pascual, M. Network structure, predator–prey modules, and stability in large food webs. *Theor. Ecol.* **1**, 55–64 (2008).
17. Gross, T., Rudolf, L., Levin, S. A. & Dieckmann, U. Generalized models reveal stabilizing factors in food webs. *Science* **325**, 747–750 (2009).
18. Tao, T., Vu, V. & Krishnapur, M. Random matrices: universality of ESDs and the circular law. *Ann. Probab.* **38**, 2023–2065 (2010).
19. Sommers, H. J., Crisanti, A., Sompolinsky, H. & Stein, Y. Spectrum of large random asymmetric matrices. *Phys. Rev. Lett.* **60**, 1895–1898 (1988).
20. Cohen, J. E., Briand, F., Newman, C. M. & Palka, Z. J. *Community Food Webs: Data and Theory* (Springer, 1990).
21. Williams, R. J. & Martinez, N. D. Simple rules yield complex food webs. *Nature* **404**, 180–183 (2000).
22. Bascompte, J., Jordano, P. & Olesen, J. M. Asymmetric coevolutionary networks facilitate biodiversity maintenance. *Science* **312**, 431–433 (2006).
23. Kokkoris, G. D., Jansen, V. A. A., Loreau, M. & Troumbis, A. Y. Variability in interaction strength and implications for biodiversity. *J. Anim. Ecol.* **71**, 362–371 (2002).
24. Wootton, J. T. & Emmerson, M. Measurement of interaction strength in nature. *Annu. Rev. Ecol. Evol. Syst.* **36**, 419–444 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Bergelson, L.-F. Bersier, A. M. de Roos, A. Eklof, C. A. Klausmeier, S. P. Lalley, R. M. May, K. S. McCann, M. Novak, P. P. A. Stanciczenko and J. D. Yeakel for comments and discussion. This research was supported by National Science Foundation grant EF0827493.

Author Contributions All authors contributed equally.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.A. (sallesina@uchicago.edu).

Diverse transitional giant fleas from the Mesozoic era of China

Diying Huang¹, Michael S. Engel^{2,3}, Chenyang Cai¹, Hao Wu¹ & André Nel⁴

Fleas are one of the major lineages of ectoparasitic insects and are now highly specialized for feeding on the blood of birds or mammals¹. This has isolated them among holometabolous insect orders, although they derive from the Antliophora (scorpionflies and true flies). Like most ectoparasitic lineages, their fossil record is meagre and confined to Cenozoic-era representatives of modern families¹, so that we lack evidence of the origins of fleas in the Mesozoic era. The origins of the first recognized Cretaceous stem-group flea, *Tarwinia*, remains highly controversial¹. Here we report fossils of the oldest definitive fleas—giant forms from the Middle Jurassic and Early Cretaceous periods of China. They exhibit many defining features of fleas but retain primitive traits such as non-jumping hindlegs. More importantly, all have stout and elongate sucking siphons for piercing the hides of their hosts, implying that these fleas may be rooted among the pollinating ‘long siphonate’ scorpionflies of the Mesozoic. Their special morphology suggests that their earliest hosts were hairy or feathered ‘reptilians’, and that they radiated to mammalian and bird hosts later in the Cenozoic.

The studied material (nine specimens) belongs to two different genera, one with two species from the Middle Jurassic Jiulongshan Formation at Daohugou (approximately 165 million years (Myr) ago), Ningcheng County, Inner Mongolia; and one monotypic genus from the Lower Cretaceous Yixian Formation (approximately 125 Myr old), at Huangbanjigou, Beipiao City, Liaoning Province. These fleas are extremely rare among the approximately 100,000 fossil insects in the collections of the Nanjing Institute of Geology and Palaeontology (Chinese Academy of Sciences). Males and females are known for the two genera. Taxonomic descriptions will be published elsewhere.

Several characters are shared by these fleas. These wingless insects, with a dorsoventrally compressed body, are distinctly larger than Recent fleas, with body lengths of 14.0–20.6 mm in females (Fig. 1a, g, Supplementary Fig. 2a and Supplementary Information 1), 8.0–14.7 mm in males (Fig. 1e, Supplementary Fig. 1a, b and Supplementary Table 1). The females from the Middle Jurassic period bear several setae and scattered ctenidia on the body, particularly on the legs (Figs 1a, b, g, 2a–c, 3a, d and Supplementary Figs 1d, h–j, 2a, c, f–h, 3b, c, f), whereas modern Siphonaptera lack ctenidia on the metatibiae. The most impressive character is their long and serrate piercing-suctorial siphon, longer in the female (Figs 1a, d, g, i, 2b, Supplementary Figs 1g and 2a, d, e) than in the male (Figs 1f, 2c, 3a, b and Supplementary Fig. 1a, f). The abdomens are large and broad with sclerites that are more reduced in females than in males (probably an adaptation for absorbing liquids, as in modern bloodfeeders) (Figs 1a, e, g, 3a, d, Supplementary Figs 1a, b, 2a, c and 3e). The highly modified pretarsal claws are long, nearly perpendicular to the last tarsal segment, and slightly curved with a prominent basal lobe forming a deep notch (Fig. 1f, Supplementary Figs 1e and 2f–h), a specialization for avoiding entanglement among hairs or feathers. A more complete list of characters is given in the Supplementary Information.

These new fossils have a large and broad abdomen, a relatively small head and thorax, long legs, no wings, short and compact antenna, ctenidia on legs, probably long siphon-like mouthparts and male genitalia that are exposed and large¹. This general body shape is shared with the younger, Aptian-aged *Tarwinia* flea from Australia^{1–3}. Unlike our fossils, *Tarwinia* has a 15-segmented flagellum³ instead of 14- to 17-segmented (Fig. 1c, e, h, Supplementary Figs 1c, 2b and 3a, d) distinctive metatibial ctenidia¹, and lacks stout, posteriorly directed setae. *Saurophthirus*, a genus of flea from the Early Cretaceous of Russian Transbaikalia, is also a parasitic insect that may have had siphonate mouthparts, compact antenna and posteriorly directed setae¹. It probably has very close affinity with our fleas from the Mesozoic era.

The most notable structure of these new fossils is their long siphonate mouthparts that are comprised of two sealed half-tubes of labial origin with segmentation that is apparently lost and a pair of very long lacinia (Figs 1d, i, 3b, Supplementary Figs 1g and 2a, d, e). These fossils differ from all of the paraneopteran orders in their five-segmented tarsi (Figs 2a–c, 3a and Supplementary Fig. 2f–h). The large and well-developed male genitalic structures (Figs 1e, 2c, 3a, c and Supplementary Fig. 1a, b) are reminiscent of the similarly prominent structures of many Antliophora and Amphismenoptera^{1,4}. These structures are generally less developed in other Holometabola, and those of the modern fleas are reduced still further and do not project so markedly from the apex of the abdomen^{5,6}. These Mesozoic insects differ from the Mecoptera in a set of characters that is shared by the modern fleas (Supplementary Fig. 4). These structures are correlated with haematophagy, and are present in several other insect groups such as the Phthiraptera (lice) and even among some true flies.

Recent and Cenozoic fleas have puncturing, suctorial and posteriorly directed mouthparts without mandibles, long and sharp half-tube labial palps and long serrate lacinia, homologous to the mouthparts of the new fossils^{5,7}. In the modern fleas the labial palps can be incompletely divided into segments, a feature that may correspond to the one-segmented labial palps in our fossils (Fig. 1d, i, Supplementary Fig. 1f and 2e). An apparent difference is that the mouthparts of modern fleas are relatively short compared to those of the Mesozoic fleas, but because they can extend as far as the apex of the thorax, they are proportionally about the same length relative to the overall body size, because the body structures of modern fleas are more compact. The mouthpart structures of these Mesozoic fleas closely resemble those of the Mesozoic siphonate mecopteran family Aneuretopsychidae (and probably those of the Mesopsychidae), except for their prognathous position. Even the ornamentation of the laciniae (bearing small teeth) is identical between the Mesozoic fleas (Fig. 1d, i and Supplementary Fig. 2e) and the Aneuretopsychidae⁸. Therefore, it is likely that the Mesozoic flea lineage is rooted near the nectar-feeding siphonate Mecoptera, if the Mesozoic fleas acquired new specializations in relation to their adaptation to haematophagy (listed in legend of Fig. 4).

¹State Key Laboratory of Palaeobiology and Stratigraphy, Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences, Nanjing 210008, China. ²Division of Entomology, Natural History Museum, University of Kansas, Lawrence, Kansas 66049-2811, USA. ³Department of Ecology & Evolutionary Biology, University of Kansas, Lawrence, Kansas 66049-2811, USA. ⁴CNRS UMR 7205, CP 50, Entomologie, Muséum National d'Histoire Naturelle, Paris F-75005, France.

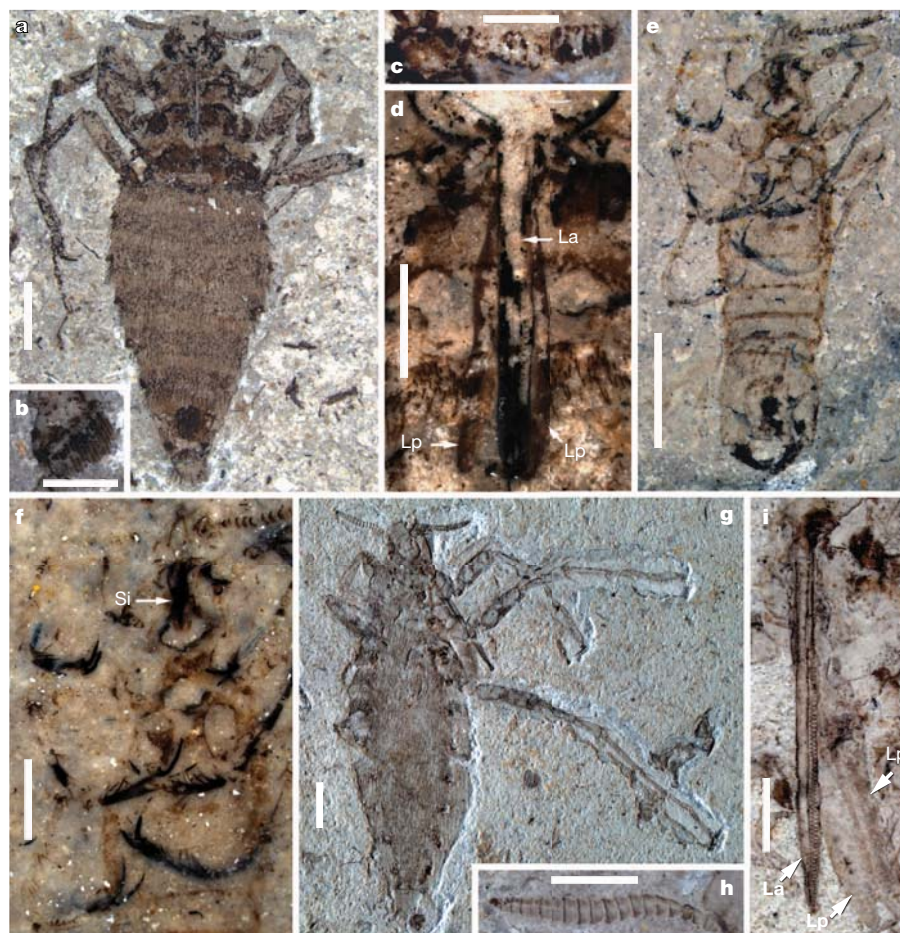


Figure 1 | Giant fleas from the Middle Jurassic of Daohugou, China.

a–f, Taxon A. **g–i**, Taxon B. **a**, 154244a, female, general habitus. **b**, Enlargement of **a**, mesotibial ctenidia. **c**, Enlargement of **a**, antenna. **d**, Mouthparts of 154244b. **e**, 154245, male, general habitus. **f**, Enlargement of **e**, mouthparts, fifth

tarsomeres and claws. **g**, 154247a, female, general habitus. **h**, Enlargement of **g**, antenna. **i**, Mouthparts of 154247a. Scale bars, 2 mm in **a**, **e** and **g**, 1 mm in **d**, **f**, **h** and **i**, 500 μ m in **b** and **c**. Gc, gonocoxite; Gs, gonostylus; La, laciniae; Lp, labial palps; Si, siphon.

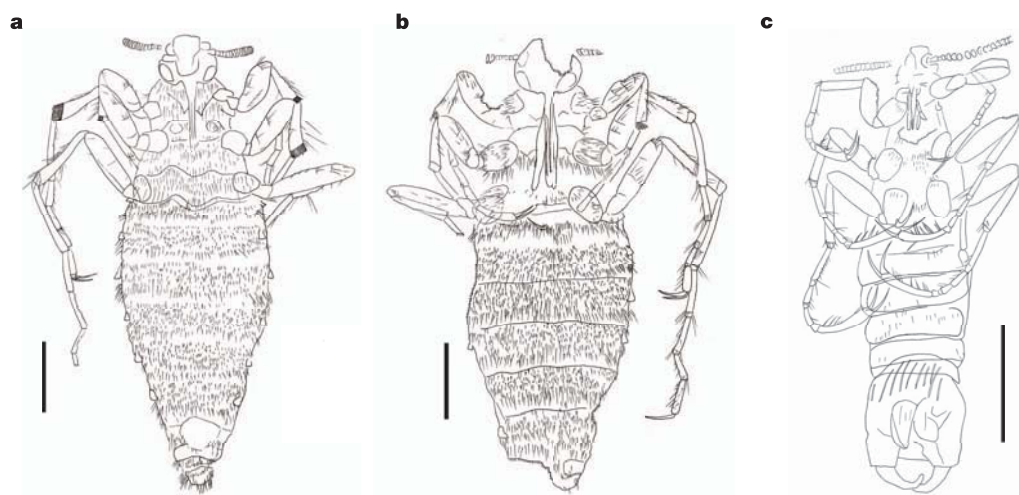


Figure 2 | Line drawings of the Middle Jurassic period fleas of Taxon A. **a**, 154244a, female, imprint. **b**, 154244b, counter-imprint of **a**. **c**, 154245, male. Scale bars, 2 mm.

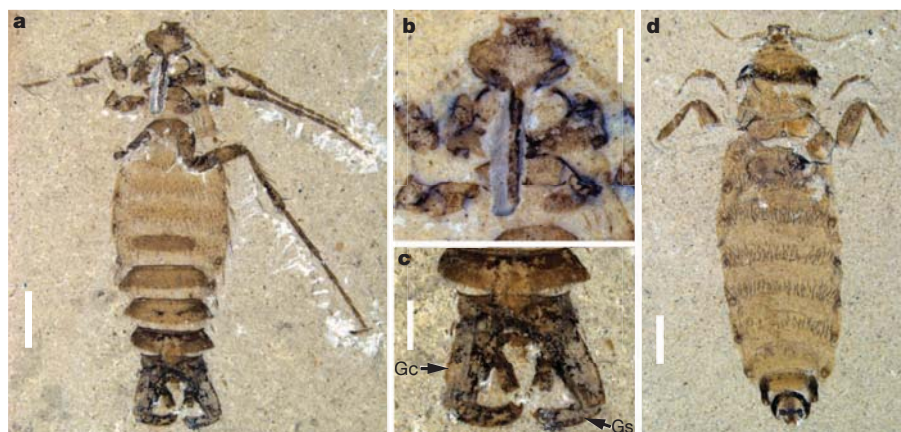


Figure 3 | Giant fleas from the Early Cretaceous period of Huangbanjigou. a–d, Taxon C. a, 154249a, male, general habitus. b, Enlargement of a, mouthparts. c, Enlargement of a, male genitalia. d, 154250a, female, general habitus. Scale bars, 2 mm in a and d, 1 mm in b and c.

Molecular studies have suggested that the scorpionflies are paraphyletic, as traditionally defined, and this means that the Boreidae would be the living sister-group to the true fleas^{1,9,10}. However, a recent work suggests that the Boreidae are possibly the sister group of the Siphonaptera plus Diptera¹¹, although this clade is weakly supported by characters that are not visible in our fossils. Nevertheless, the Boreidae have strong mandibles, relatively short lacinia and labial

palps^{4,12}, all of which are plesiomorphies that do not correspond to the mouthpart structures of the modern fleas and to those of Mesozoic fleas. A dipteran–nannomecopteran–siphonapteran clade rather than a Boreidae–dipteran–siphonapteran clade was recently supported¹³. The Nannochoristidae have tube-like labial palps with elongate lacinia and very small mandibles¹⁴, and these resemble the siphon structures of the siphonate Mecoptera, the new fossils and the modern fleas. The Nannochoristidae are currently considered “liquid feeders, most likely on nectar of flowers and/or juices of fruits” (ref. 15), which corresponds well to the Mesozoic siphonate Mecoptera. Hence, the affinities of the fleas and flies within the mecopteran ‘stock’ remain uncertain. Furthermore, the molecular phylogenetic studies^{9,10} ignore several fossil groups and particularly the extinct ‘long siphonate mecopteran’ families. Thus, the results of the molecular analyses do not a priori conflict with possible close affinities between the siphonate mecopteran lineage, our Mesozoic species and the modern fleas.

The idea of a fleas–flies sister-group relationship implies that the two lineages that gave rise to the Diptera and the Siphonaptera separated during the Permian period. It also suggests that one lineage evolved early to become the Diptera during the Early Triassic period¹ and that the second diversified into the nectar-sucking siphonate Mecoptera and the blood-sucking fleas during the Jurassic with a further specialization into the modern fleas during the Mesozoic and Palaeogene periods. It seems that the arrival of fleas in the Jurassic was delayed compared to the arrival of flies in the Triassic, perhaps because the fleas needed vertebrate hosts with fur or feathers that emerged in the Late Triassic.

The alate siphonate Mecoptera are thought to have fed on “ovular secretions of extinct gymnosperms” (ref. 15). The switch by the insects from feeding on plant fluids to haematophagy is paralleled in reverse by the haematophagous Ceratopogonidae, a sister group to the Chironomidae, which were first haematophagous, becoming nectarivorous during the Cretaceous¹⁶.

Our Mesozoic giant fleas have dorsoventrally flattened bodies and unmodified metacoxae and metafemora. This is different from those features of modern adult fleas that facilitate movements among the feathers or fur of their hosts; hindlegs with enlarged coxae and femora that are modified for jumping, and laterally compressed bodies. Thus, these Mesozoic species and modern fleas have very different chorologies. Although their larvae are generally detritivorous in the nests of their hosts, the adults of nearly all modern fleas are ectoparasites of mammals, with multiple derived lineages shifting to birds in several higher families¹⁷. The most basal family of modern fleas (Tungidae) principally feed on sloths and armadillos, whereas other basal families attack metatherians and rodent lineages. These associations have supported the idea that fleas diversified at the same time as the Theria, and as long ago as the Late Jurassic¹⁷, correlating well with

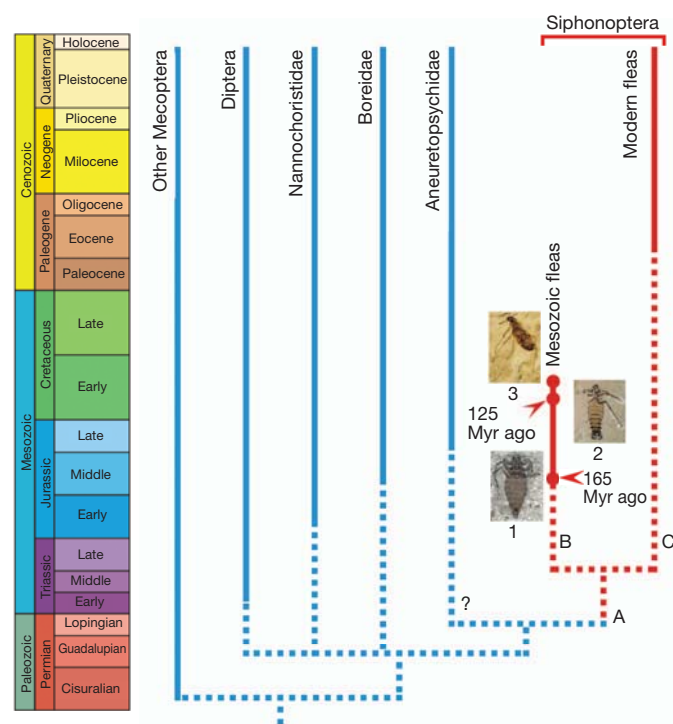


Figure 4 | Phylogeny of Siphonaptera and related clades. The clade that includes Aneuretopsyichidae and Siphonaptera is supported by the presence of a stylate proboscis with serrate laciniae. Clade A (Siphonaptera) is characterized by a complete absence of wings, loss of or greatly reduced eyes, absence of ocelli, a compact antennal flagellum, posteriorly directed setae and distinctive ctenidia, and long pretarsal claws with a prominent basal lobe. Clade B (Mesozoic fleas) is characterized by the presence of ctenidia on the tibiae, a dorsoventrally compressed body, and largely desclerotized and membranous abdominal sternites in females. Clade C (modern fleas) is characterized by a laterally flattened body, saltatorial hindlegs, eyes that are highly reduced, subcuticular or lost, and highly reduced external portions of the male genitalia (gonocoxae, gonostyli). 1, Middle Jurassic fleas; 2, Early Cretaceous fleas; 3, *Tarwinia*.

the age of our fossils. Indeed, it is possible that our Mesozoic fleas were ectoparasites of early therians or other early mammals such as multituberculates, which were diverse and abundant at the time, and such taxa are well represented in the Jurassic and Early Cretaceous of China at the same or adjacent localities^{18–21}.

However, the early mammals were small animals, making the large size of these Mesozoic species and the robustness of their mouthparts seem mismatched. The stiff, elongate, posteriorly directed setae and the long curved claws with a prominent basal lobe are clearly specialized structures in which hairs may become firmly lodged. This indicates that these Mesozoic species were living on hosts that were covered by fur or feathers, rather than on vertebrates with skin that was nude or covered with scales. Thus, it is also possible that the hosts of these early fleas were among the feathered dinosaurs of the period that became well-known from the same deposits^{18,22}. *Saurophthirus* was purportedly a pterosaur ectoparasite²³. Its very elongated and slender legs suggest that it probably lived on very different animals, indicating diversity among Mesozoic vertebrate ectoparasites.

Alternatively, it is possible that these early fleas, in contrast to the adults of their modern counterparts, did not live exclusively on their host, like some modern earwigs that are epizoid on bats, but fed on a range of vertebrate lineages, including early mammals. This would suggest that these early fleas eventually shifted from non-mammalian hosts to the Theria and then became specialized further, perhaps late in the Cretaceous period or in the earliest Cenozoic era. The much later diversification of placental mammals (Eutheria) supported the substantial radiation of higher Eusiphonaptera and particularly of the diverse Neosiphonaptera (Ceratophyllomorpha and Pulicomorpha), eventually giving rise to the plague-carrying species that have so markedly altered the course of human history.

METHODS SUMMARY

The material is housed at Nanjing Institute of Geology and Palaeontology. The fossils were prepared using a sharp knife. Photographs were taken using the Zeiss Discovery V20 microscope system, moistened with 70% alcohol (Figs 1c, d, f, 3b, Supplementary Figs 1a, d–j, 2b–d and 3a–d, f) and drawings were made using a camera lucida on a binocular Olympus SZX7.

Received 22 November 2011; accepted 6 January 2012.

Published online 29 February 2012.

1. Grimaldi, D. & Engel, M. S. *Evolution of the Insects* (Cambridge Univ. Press, 2005).
2. Riek, E. F. Lower Cretaceous fleas. *Nature* **227**, 746–747 (1970).
3. Jell, P. A. & Duncan, P. M. Invertebrates, mainly insects, from the freshwater Lower Cretaceous, Koonwarra Fossil Bed (Korumburra Group), South Gippsland, Victoria. *Mem. Assoc. Australas. Palaeontol.* **3**, 111–205 (1986).
4. Penny, N. D. A systematic study of the family Boreidae (Mecoptera). *Univ. Kansas Sci. Bull.* **51**, 141–217 (1977).
5. Snodgrass, R. E. The skeletal anatomy of fleas (Siphonaptera). *Smith. Misc. Coll.* **104**, 1–89 (1946).

6. Cheetham, T. B. *Male Genitalia and Phylogeny of Pulicoidea* (Koeltz Sci. Press, 1988).
7. Kluge, N. J. The homology of mouthparts in fleas (Insecta, Aphaniptera). *Entomol. Rev.* [transl.] **82**, 1020–1026 (2002).
8. Ren, D. et al. A probable pollination mode before angiosperms: Eurasian, longproboscis scorpionflies. *Science* **326**, 840–847 (2009).
9. Whiting, M. F., Carpenter, J. M., Wheeler, Q. D. & Wheeler, W. C. The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal sequences and morphology. *Syst. Biol.* **46**, 1–68 (1997).
10. Whiting, M. F. Mecoptera is paraphyletic: multiple genes and phylogeny of Mecoptera and Siphonaptera. *Zool. Scr.* **31**, 93–104 (2002).
11. Friedrich, F. & Beutel, R. G. The thoracic morphology of *Nannochorista* (Nannochoristidae) and its implications for the phylogeny of Mecoptera and Antliophora. *J. Zool. Syst. Evol. Res.* **48**, 50–74 (2010).
12. Beutel, R. G., Friedrich, F. & Whiting, M. F. Head morphology of *Caurinus* (Boreidae, Mecoptera) and its phylogenetic implications. *Arthropod Struct. Dev.* **37**, 418–433 (2008).
13. Schneeberg, K. & Beutel, R. G. The adult head structures of Tipulomorpha (Diptera, Insecta) and their phylogenetic implications. *Acta Zool.* **92**, 316–343 (2011).
14. Beutel, R. G. & Baum, E. Longstanding entomological problem finally solved? Head morphology of *Nannochorista* (Mecoptera, Insecta) and possible phylogenetic implications. *J. Zoological Syst. Evol. Res.* **46**, 346–367 (2008).
15. Palmer, C. M. Diversity of feeding strategies in adult Mecoptera. *Terrestr. Arthr. Rev.* **3**, 111–128 (2010).
16. Azar, D., Veltz, I. & Nel, A. Mandibulate chironomids: primitive or derived? (Diptera: Chironomidae). *Syst. Entomol.* **33**, 688–699 (2008).
17. Whiting, M. F., Whiting, A. S., Hastriter, M. W. & Dittmar, K. A molecular phylogeny of fleas (Insecta: Siphonaptera): origins and host associations. *Cladistics* **24**, 677–707 (2008).
18. Benton, M. J., Zhang, Z., Orr, P. J., Zhang, F. & Kearns, S. L. The remarkable fossils from the Early Cretaceous Jehol Biota of China and how they have changed our knowledge of Mesozoic life. *Proc. Geol. Assoc.* **119**, 209–228 (2008).
19. Luo, Z.-X., Ji, Q. & Yuan, C. Convergent dental adaptations in pseudo-tribosphenic and tribosphenic mammals. *Nature* **450**, 93–97 (2007).
20. Meng, J., Hu, Y., Wang, Y., Wang, X. & Li, C. A Mesozoic gliding mammal from northeastern China. *Nature* **444**, 889–893 (2006).
21. Luo, Z.-X., Yuan, C., Meng, Q. & Ji, Q. A Jurassic eutherian mammal and divergence of marsupials and placentals. *Nature* **476**, 442–445 (2011).
22. Zhang, F., Zhou, Z., Xu, X., Wang, X. & Sullivan, C. A bizarre Jurassic maniraptoran from China with elongate ribbon-like feathers. *Nature* **455**, 1105–1108 (2008).
23. Ponomarenko, A. G. A new insect from the Cretaceous of Transbaikalia, a possible parasite of pterosaurs. *Paleontol. J.* **10**, 339–343 (1976).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to X. D. Wang for presenting specimen 154244, A. Short, X. Xu, F. C. Zhang, and X. L. Wang for discussions and S. Davis and X. Y. Fan for technical assistance. Financial support was provided by the National Natural Science Foundation of China, National Basic Research Program of China (2012CB821900), Chinese Academy of Sciences (KZCX2-YW-QN104) and the US National Science Foundation (DEB-0542909).

Author Contributions All authors participated in morphological studies. D.H., M.S.E. and A.N. prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.H. (huangdiying@sina.com) or A.N. (anel@mnhn.fr).

Treatment of stroke with a PSD-95 inhibitor in the gyrencephalic primate brain

Douglas J. Cook¹, Lucy Teves¹ & Michael Tymianski^{1,2,3,4}

All attempts at treating strokes by pharmacologically reducing the human brain's vulnerability to ischaemia have failed, leaving stroke as a leading cause of death, disability and massive socio-economic loss worldwide¹. Over decades, research has failed to translate over 1,000 experimental treatments from discovery in cells and rodents to use in humans^{2–4}, a scientific crisis that gave rise to the prevailing belief that pharmacological neuroprotection is not feasible or practicable in higher-order brains. To provide a strategy for advancing stroke therapy, we used higher-order gyrencephalic non-human primates, which bear genetic, anatomical and behavioural similarities to humans^{5,6} and tested neuroprotection by PSD-95 inhibitors—promising compounds that uncouple postsynaptic density protein PSD-95 from neurotoxic signalling pathways^{7–10}. Here we show that stroke damage can be prevented in non-human primates in which a PSD-95 inhibitor is administered after stroke onset in clinically relevant situations. This treatment reduced infarct volumes as gauged by magnetic resonance imaging and histology, preserved the capacity of ischaemic cells to maintain gene transcription in genome-wide screens of ischaemic brain tissue, and significantly preserved neurological function in neurobehavioural assays. The degree of tissue neuroprotection by magnetic resonance imaging corresponded strongly to the preservation of neurological function, supporting the intuitive but unproven dictum that integrity of brain tissue can reflect functional outcome. Our findings establish that tissue neuroprotection and improved functional outcome after stroke is unequivocally achievable in gyrencephalic non-human primates treated with PSD-95 inhibitors. Efforts must ensue to translate these findings to humans.

All stroke neuroprotectants have failed to be translated to clinical human use³, though many were seemingly effective in rodents. To bridge this biological gap, we conducted experiments in cynomolgus macaques. These non-human primates (NHPs) possess higher-order brains similar to humans and a rich behavioural repertoire amenable to assessments using standardized tests¹¹ (see Supplementary Discussion for choice of primate species). They are amenable to magnetic resonance imaging (MRI), and to the quantification of the ischaemic penumbra, the portion of ischaemic brain that is potentially salvageable if an appropriate treatment is given¹².

Although the MRI definition of a penumbra is controversial^{12,13}, we operationally defined it as that brain region in which the perfusion- and diffusion-weighted MR images are mismatched (PWI/DWI mismatch). We then developed two experimental models of surgical middle cerebral artery occlusion (MCAO) in the NHPs in which the PWI/DWI mismatch evolves at different rates (Supplementary Methods). In the most severe model, the MCAO is performed distally to the orbito-frontal branch of the MCA (Supplementary Fig. 1a, b, d, f). This isolates the ischaemic brain from the collateral circulation to create a large, rapidly evolving stroke with a rapidly shrinking penumbra such that, by 3 h, more than 80% of the MRI perfusion defect exhibits a DWI signal (MCAO⁽⁺⁾P; Supplementary Fig. 1g, lower panels). In the second model, the MCAO is proximal to the

orbito-frontal branch, allowing a small collateral circulation to slow the evolution of the DWI signal such that, at 3 h after MCAO, less than 30% of the perfusion defect exhibits a DWI signal, signifying a stroke with an ischaemic penumbra (MCAO⁽⁺⁾P; Supplementary Fig. 1c, e). However, as in humans, this penumbra evolves to a completed stroke in the absence of treatment (Supplementary Fig. 1g, upper panels).

To test the feasibility of neuroprotection in the high-order brain we used the 20-mer peptide termed Tat-NR2B9c, comprising the nine carboxy-terminal amino acids of the *N*-methyl-D-aspartate receptor (NMDAR) NR2B subunit fused to the 11-mer HIV-1 Tat protein transduction domain⁷. This agent meets recommendations for preclinical stroke drug development established by the Stroke Academic Industry Roundtable (STAIR) committee^{14,15}. It acts by perturbing the protein–protein interactions of PSD-95 (ref. 16), a synaptic scaffolding protein that links NMDARs to neurotoxic signalling pathways^{8,17}. Treating rats subjected to MCAO with Tat-NR2B9c is neuroprotective in multiple rodent stroke models^{7,10,18}, making it a promising candidate for testing in NHPs. Our overall rationale for using Tat-NR2B9c is further described in the Supplementary Discussion and Supplementary Table 1.

We first evaluated Tat-NR2B9c in NHPs subjected to MCAO⁽⁺⁾P, rationalizing that efficacy in this stroke model, though it produces more severe strokes than are usually seen in humans, might maximize the chance of ultimate clinical use. The primary outcome measure was infarct volume at 30 days measured from a *T*₂-weighted MRI study. Anatomical secondary outcomes were infarct volumes at 4 h and 24 h by DWI MRI, at 24 h by *T*₂ MRI and at 30 days by *T*₂ MRI and histology (Supplementary Fig. 2a). Neurobehavioural outcomes were measured throughout the 30-day observation period using the non-human primate stroke scale (NHPSS)¹¹ and a sensorimotor battery of tasks comprising the hill and valley task, two-tube task and six-well task (Supplementary Fig. 2b, c and Supplementary Methods)¹⁹.

Twenty macaques were randomized to receive a 10-min intravenous infusion of Tat-NR2B9c (2.6 mg kg^{−1}) or placebo (0.9% saline) beginning 1 h after the onset of a 90-min MCAO⁽⁺⁾P (Supplementary Methods and Supplementary Fig. 2a). The dose selected for NHPs was approximated from calculations of a 'primate equivalent dose' extrapolated from previous doses used in rat studies¹⁰ and was based on normalization to interspecies differences in body surface area²⁰.

Animals were transferred to the MRI scanner within 15 min of MCAO and all underwent perfusion imaging to quantify the brain volume deprived of blood flow during MCAO (tissue-at-risk¹³). Additionally, MR angiography was conducted to confirm MCAO (Supplementary Methods and Supplementary Fig. 2a). A second MR angiography was performed to confirm reperfusion after the 90-min MCAO, followed by diffusion imaging at 4 h. Animals were then awakened and allowed to recover. They were re-anaesthetized and re-imaged at 24 h and at 30 days (Supplementary Fig. 2a). NHPSS scores were assigned within 8 h of MCAO and up to 30 days, and the remaining neurological tests were conducted on days 7 and 30 (Supplementary Fig. 2b, c).

¹Toronto Western Hospital Research Institute, Toronto, Ontario, M5T 2S8, Canada. ²Department of Physiology, University of Toronto, Toronto, Ontario, M5S 1A8, Canada. ³Institute of Medical Science, University of Toronto, Toronto, Ontario, M5S 1A8, Canada. ⁴Department of Surgery, University of Toronto, Toronto, Ontario, M5S 1A8, Canada.

Four of ten animals receiving placebo died within 48 h of their strokes owing to brain swelling and uncal herniation. Three animals treated with Tat-NR2B9c died as a result of surgical/anaesthetic complications unrelated to stroke or to drug (Supplementary Discussion). None were excluded from the 'intent-to-treat' analysis (see Supplementary Discussion for impact of the intent-to-treat approach). All missing data due to early mortalities were imputed to reflect the largest possible infarct volumes and worst neurological scores. Although this approach biases against detecting a significant treatment effect, it is the most conservative, and the most reflective of that used in human clinical trials.

There were no differences between the drug and placebo groups in physiological parameters (Supplementary Table 2a) or in the volume of tissue at risk as determined by perfusion imaging within 15 min of MCAO (Fig. 1a). However, by 24 h, animals treated with Tat-NR2B9c exhibited a significant reduction in infarct volume compared with placebo by DWI imaging (44.0% reduction; $P = 0.039$; Fig. 1b) and by T_2 -weighted imaging (37.4% reduction; $P = 0.010$; Fig. 1b, c). This reduction in infarct volume persisted as reflected by the 30-day T_2 -weighed MRI scans (38.7% reduction; $P = 0.013$; Fig. 1c) and by histological evaluation at 30 days (Fig. 1d; 59.3% and 73.6% reduction in infarct volume when evaluated by intent-to-treat and with early mortalities removed, respectively; $P < 0.001$). Because NHPs, like humans, may have variable infarcts after MCAO, the infarct volume of each animal was normalized to its MRI perfusion defect measured within 15 min of MCAO. This normalization revealed that treatment with Tat-NR2B9c reduced infarcts by 55% of the volume at risk by 24 h as gauged by DWI imaging, and by 70% at 30 days as measured with T_2 -weighed MRI (Fig. 1e). Infarct volumes calculated from the 24 h DWI MRI correlated well with those obtained from the histological analysis at 30 days ($R = 0.691$, $P < 0.01$).

We conducted neurological assessments throughout the 30-day observation period using the NHPSS¹¹ and a sensorimotor battery of tasks including the hill and valley task, two-tube choice task and six-well task¹⁹ (Supplementary Methods). The NHPSS is a composite of ratings analogous to the National Institutes of Health Stroke Scale used in human stroke trials²¹. A score of 41 points represents severe bilateral neurological impairment and 0 is normal. The remaining tests measure a combination of overall strength of the extremity, fine motor function and the influence of a hemi-neglect or visual field defect (Supplementary Methods and Supplementary Fig. 2b, c).

Animals treated with Tat-NR2B9c exhibited improved NHPSS scores from the earliest assessment at 8 h after ischaemia onset and throughout the 30-day observation period ($P = 0.018$, two-way repeated-measures analysis of variance (ANOVA); Fig. 1f). Performance in the two-tube choice task returned to pre-stroke levels in animals treated with Tat-NR2B9c, but remained completely impaired in the placebo group (Supplementary Fig. 3a), suggesting that brain salvage prevented 'extinction', the tendency for attention to items in ipsilesional hemisphere to overshadow attention to items in contralateral hemisphere¹⁹. Treatment with Tat-NR2B9c also significantly improved the performance of animals in the six-well (Supplementary Fig. 3b) and the hill and valley staircase tasks in the left upper extremity (Supplementary Fig. 3c, d). Right upper extremity performance also showed improvements, suggesting overall improved attention and perceptual ability (Supplementary Fig. 3e, f).

In human studies, CT or MRI measurements of infarct volumes have demonstrated no or modest correlations with clinical outcomes^{22,23}. Thus, the intuitive dictum that infarct size reduction is beneficial is unproven. In the current study, the 24-h DWI and 30-day T_2 -weighed MRI correlated highly with the 30-day NHPSS (Supplementary Fig. 3g, h). This suggests that infarct volumes measured by MRI of the high-order brain predict neurological performance after stroke, and may comprise a reasonable surrogate measure of clinical outcome.

Cerebral ischaemia has long been known to inhibit global protein synthesis and gene transcription profoundly²⁴, suggesting that impaired

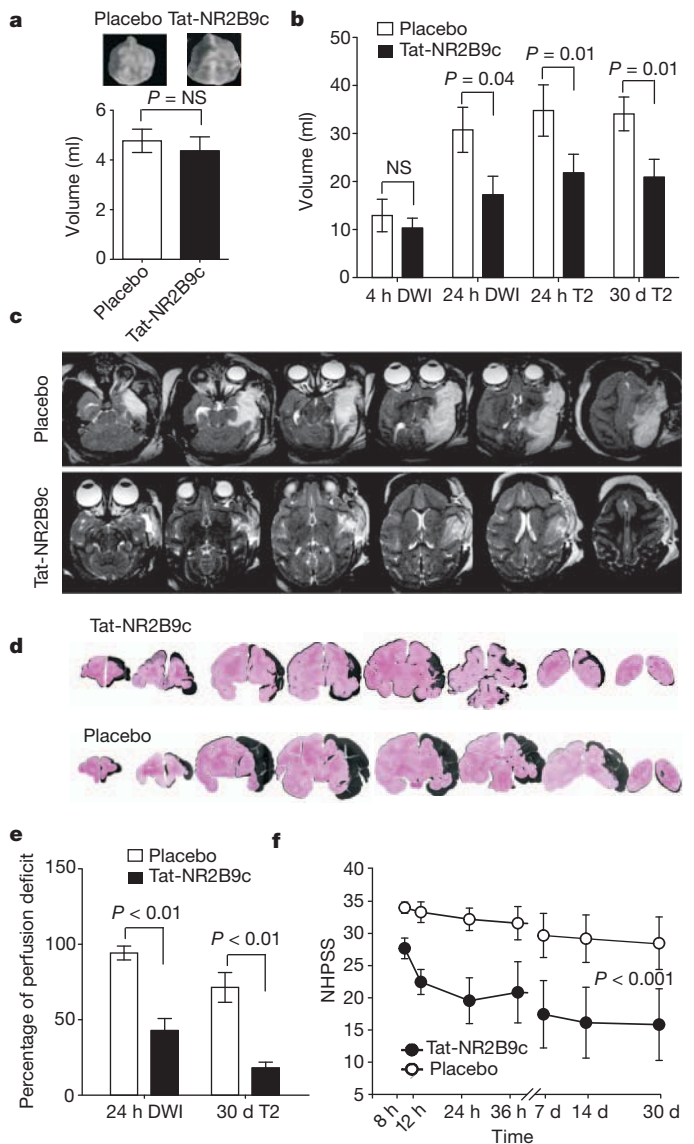


Figure 1 | Treatment with Tat-NR2B9c attenuates infarct volume in NHPs subjected to MCAO^(-P). **a**, Volumes of perfusion defects (Methods) at baseline. NS, not significant. **b**, Analysis of stroke volumes as measured by DWI and T_2 MRI over 30 days. **c**, Representative T_2 -weighted images of strokes incurred in placebo and drug-treated animals 24 h after MCAO^(-P). **d**, Representative serial histological sections were stained with haematoxylin and eosin taken 30 days after MCAO^(-P) from Tat-NR2B9c- and placebo-treated animals. Shaded areas represent the anticipated brain area based on mirroring the contralateral side. **e**, Stroke volumes calculated using 24-h DWI and 30-day T_2 -weighted MRI scans normalized to the initial perfusion defects in each animal. **f**, NHPSS over the 30-day observation period. Symbols, bars and error bars in **a**, **b**, **e**, **f** indicate means of $n = 10$ per group \pm s.e.m.

transcriptional capacity is a measure of dysfunction in affected cells. To examine whether neuroprotection by Tat-NR2B9c may prevent such dysfunction, we examined the transcription response on a genome-wide level in macaques subjected to stroke. Six additional macaques were subjected to a permanent MCAO^(-P), and were treated either with Tat-NR2B9c or with placebo beginning at 5 min after ischaemia onset. The animals were placed in the MRI scanner, and DWI MRI scans were obtained every 15 min. The volume of brain in which DWI hyperintensity was detectable increased over time in both groups. However, treatment with Tat-NR2B9c attenuated the rate of this increase by about twofold (Fig. 2a; time constants = 2.20 ± 0.28 h and 4.50 ± 0.54 for control and Tat-NR2B9c, respectively; $P = 0.019$). Moreover, within the ischaemic volume, the DWI intensity in brains of

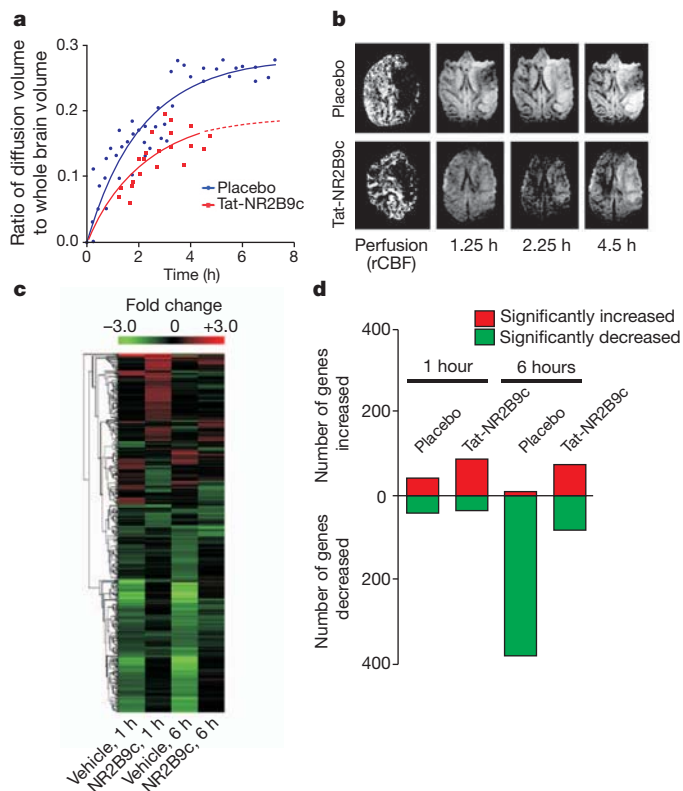


Figure 2 | Treatment with Tat-NR2B9c preserves capacity for transcription in NHPs subjected to stroke. Animals were subjected to MCAO⁽⁻⁾P and treated within 5 min with Tat-NR2B9c ($n = 3$) or placebo ($n = 3$). **a**, Time course of increase in DWI hyperintensity after MCAO⁽⁻⁾P in treated and control animals. **b**, Representative DWI images from **a**. **c**, Hierarchical cluster analysis of whole-genome response to MCAO⁽⁻⁾P from ischaemic penumbras of control and Tat-NR2B9c-treated animals. **d**, Plot of significantly differentially regulated genes in the drug- and placebo-treated NHP penumbras at 1 and 6 h after stroke.

Tat-NR2B9c-treated animals remained lower than that of untreated controls, suggesting that tissue within the infarct volume maintained better integrity (Fig. 2a, b).

We collected tissue from the ischaemic penumbras (adjacent to infarcted tissue) and from corresponding sites in the contralateral, non-ischaemic hemisphere at 1 h and 6 h after ischaemia onset (Supplementary Methods and Supplementary Fig. 4). We selected these after-stroke intervals as they correspond to timeframes during which most neuroprotectants are administered. RNA was extracted and hybridized to whole-genome macaque arrays (Supplementary Methods). Differential gene expression analysis revealed that treatment with Tat-NR2B9c resulted in a lower proportion of downregulated genes and preservation of the capacity to upregulate genes at both 1 h and 6 h after stroke (Fig. 2c, d and Supplementary Table 3). Specific genes and pathways that were differentially regulated in placebo and drug-treated animals included several that are associated with endogenous cytoprotective responses to ischaemia and cell stress (Supplementary Data, Supplementary Fig. 5 and Supplementary Tables 3 and 4). Overall, this genome-wide survey suggests that neuroprotection with Tat-NR2B9c preserved cellular functionality as gauged by the capacity for gene transcription in ischaemic brain tissue.

Our experiments demonstrate the feasibility of neuroprotection in the higher-order brain when treatment with a PSD-95 inhibitor is initiated 60 min after the onset of a 90-min MCAO⁽⁻⁾P. However, only a small proportion of stroke victims may fit this pattern. Consequently, we next evaluated the use of treatment with Tat-NR2B9c in experiments having a broader clinical applicability. Currently, the only widely approved treatment for acute ischaemic stroke is reperfusion

of occluded brain arteries using the intravenous infusion of the fibrinolytic agent, recombinant tissue plasminogen activator (rt-PA). Though Tat-NR2B9c is effective in rodents in the absence of reperfusion^{10,18}, the clinical use of rt-PA suggests that building neuroprotection onto a reperfusion setting may be the most practical means to maximize its use. Reperfusion with intravenous rt-PA is most beneficial in improving clinical outcomes when administered within 90 min after stroke onset, and benefit decreases thereafter until it is marginal or nil at 4.5 h (refs 25, 26). This narrow window for the use of reperfusion limits the number of patients who might benefit. Thus one potential application of early treatment with a neuroprotectant is to extend the interval during which clinical benefit may be obtainable from reperfusion therapy. To examine this, we evaluated in the NHPs whether administration of Tat-NR2B9c 60 min after MCAO⁽⁻⁾P onset might improve stroke outcome when reperfusion is delayed until the 4.5-h time point, at which intravenous rt-PA is no longer of significant benefit in humans.

Twelve macaques were randomized to receive a 10-min intravenous infusion of Tat-NR2B9c (2.6 mg/kg) or placebo (0.9% saline) beginning 1 h after the onset of a 4.5-h MCAO⁽⁻⁾P. Otherwise, methods were similar to our first study except for the timing of MRI scans and that final imaging and neurological assessments were conducted at 7 days.

There were no mortalities and no differences between the drug and placebo groups in physiological parameters (Supplementary Table 2b) or in the volume of tissue at risk upon MCAO⁽⁻⁾P (Fig. 3a). However, despite the prolonged ischaemic interval, animals treated with Tat-NR2B9c exhibited a significant reduction in infarct volumes compared with placebo when evaluated by T_2 and DWI imaging at 48 h and by T_2 -weighted imaging at 7 days (Fig. 3b–e). Moreover, animals treated with Tat-NR2B9c exhibited improved NHPSS scores from the earliest assessment at 12 h after ischaemia onset and throughout the 7-day observation period ($P = <0.001$, two-way repeated-measures ANOVA; Fig. 3f), and trended to better performance in the six-well and the valley staircase tasks (Supplementary Fig. 6a). These results suggest that early treatment with Tat-NR2B9c may increase the window during which reperfusion may have functional benefits, even in the model of severe MCAO⁽⁻⁾P in which collateral circulation is limited and the penumbra is small (Supplementary Fig. 1). The size of the benefit of treatment at 4.5 h after stroke as gauged by MRI and by neurological evaluations suggests a potential for use of early neuroprotection to extend the benefits of reperfusion therapy even beyond the 4.5-h window.

Although treatment with a neuroprotectant within 60 min of stroke onset is feasible in a small subset of patients²⁷, extending the therapeutic window of administration would benefit a much greater proportion of stroke victims. Thus we determined whether administering Tat-NR2B9c at 3 h after stroke onset is beneficial in the setting of a prolonged MCAO. In humans, reperfusion with intravenous rtPA administered 3 h after stroke onset is beneficial even in the absence of neuroprotection²⁸. This attests to the existence of a salvageable penumbra at this time in many patients. We reproduced this clinical situation experimentally by using the MCAO⁽⁺⁾P model in which the NHPs exhibit a significant PWI/DWI mismatch (penumbra) at 3 h. Like in humans, such a mismatch progresses to infarction in the absence of treatment (Supplementary Fig. 1).

Twenty-four macaques were randomized to receive a 10-min intravenous infusion of Tat-NR2B9c (2.6 mg kg⁻¹) or placebo (0.9% saline) beginning 3 h after the onset of a 3.5 h MCAO⁽⁺⁾P. Other methods were unchanged, except that final imaging and neurological assessments were conducted at 14 days. There were no mortalities and no differences in physiological parameters (Supplementary Table 2c) or in the volume of tissue at risk upon MCAO⁽⁺⁾P between the groups (Fig. 4a). However, despite both the prolonged ischaemic interval and the delayed treatment with Tat-NR2B9c, drug-treated animals exhibited significant reductions in infarct volumes compared with placebo as evaluated by anatomical (MRI) criteria at 48 h and 14 days (Fig. 4b–e). The NHPs treated with Tat-NR2B9c also exhibited improved NHPSS

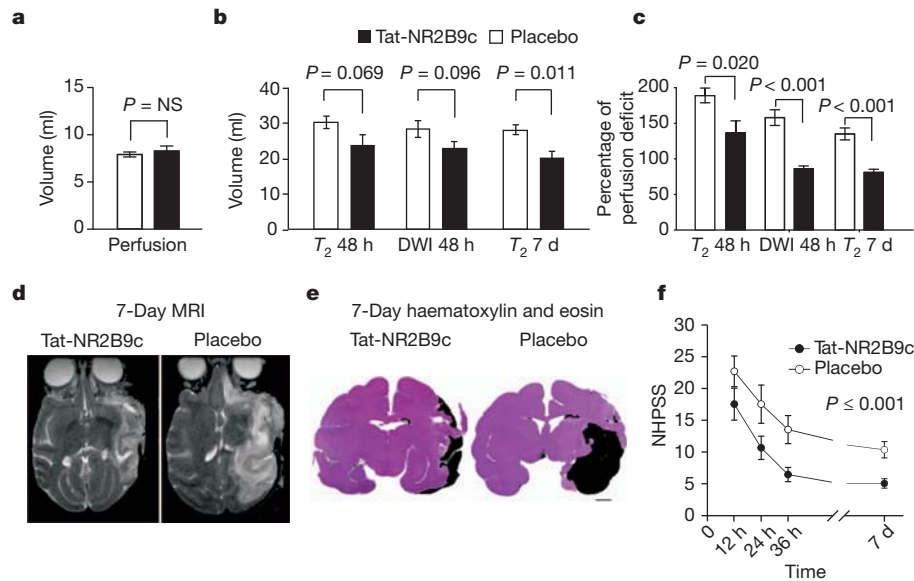


Figure 3 | Treatment with Tat-NR2B9c at 60 min improves MRI and functional outcome after a 4.5-h MCAO⁽⁺⁾P. **a**, Volumes of perfusion defects (Methods) at baseline. **b**, Stroke volumes as measured by DWI and T_2 MRI over 7 days. **c**, Stroke volumes from the 48-h DWI and T_2 - and 7-day T_2 -weighted

MRI scans normalized to each animal's initial perfusion deficit.

d, Representative 7-day MRIs. **e**, Representative 7-day histology. **f**, NHPSS scores over the 7-day observation period. Symbols, bars and error bars in **a**, **b**, **c**, **f** indicate means of $n = 6$ per group \pm s.e.m.

scores throughout the 14-day observation period ($P = 0.004$, two-way repeated-measures ANOVA; Fig. 4f), and tended to better performance in the six-well and the valley staircase tasks (Supplementary Fig. 6b). Thus, treatment with Tat-NR2B9c 3 h after stroke onset is effective in reducing stroke damage in NHPs. As this therapeutic window is practical in stroke victims²⁸, treatment with a PSD-95 inhibitor may constitute a clinically practicable therapeutic strategy.

After the failure of all past neuroprotectants in clinical trials, the feasibility of a pharmacological strategy to lessen the damaging effects of a stroke in the higher-order brain was in doubt^{4,29}. The current prevailing belief, that neuroprotection in humans is not achievable, arose in response to these failures. However, there has never been concrete evidence that stroke mechanisms in rodents and higher-order brains differ to an extent that would preclude neuroprotection in

primates or humans. Moreover, it is acknowledged that failures of clinical trials may have been related to trial design²⁹ and to shortcomings of past preclinical studies³⁰. We have provided four distinct studies illustrating relevant situations in which neuroprotection with PSD-95 inhibitors is feasible in gyrencephalic Old World primates as gauged by measures of improved functional outcomes and reduced tissue damage by MRI, by histology and by retained capacity for transcription. Our findings using PSD-95 inhibitors may not be generalizable to other neuroprotective strategies. However, they defeat the current pessimistic belief by demonstrating that pharmacological neuroprotection of the high-order brain of gyrencephalic primates is unequivocally possible. Unless there exist fundamental, as yet unknown, relevant differences between such primates and humans, neuroprotection in humans using PSD-95 inhibitors should also be feasible.

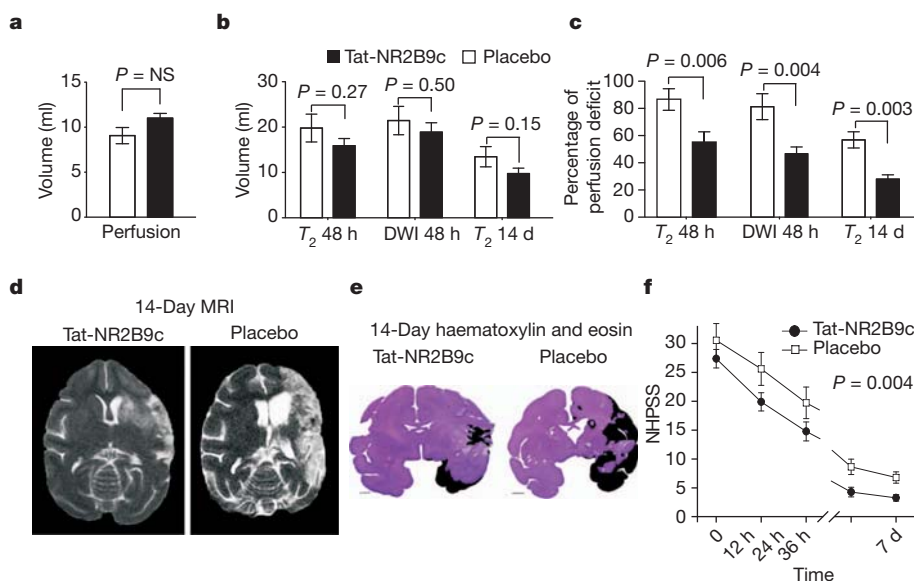


Figure 4 | Treatment with Tat-NR2B9c at 3 h improves MRI and functional outcome after a 3.5-h MCAO⁽⁺⁾P. **a**, Volumes of perfusion defects (Methods) at baseline. **b**, Stroke volumes as measured by DWI and T_2 MRI over 14 days. **c**, Stroke volumes from the 48-h DWI and T_2 and 14-day T_2 -weighted MRI

scans normalized to each animal's initial perfusion deficit. **d**, Representative 14-day MRI and **e**, representative 14-day histology. **f**, NHPSS scores over the 14-day observation period. Symbols, bars and error bars in **a**, **b**, **c**, **f** indicate means of $n = 12$ per group \pm s.e.m.

METHODS SUMMARY

All experiments were conducted in male captive-bred cynomolgus macaques (*Macaca fascicularis*; 2.4–5.0 kg). They were approved by the Animal Care Committee of the University Health Network and complied with the relevant guidelines and regulations of the Canadian Council on Animal Care and with the ‘recommendations for ensuring good scientific enquiry’ of the STAIR committee¹⁴, including allocation concealment and blinded assessment of all outcomes. MCAO was achieved under general anaesthesia through a right pterional craniotomy using a 5 mm titanium aneurysm clip. Animals were imaged on a 7 T Bruker BioSpec system running Paravision 4.0 software and using a B-GA20S gradient coil. RNA isolated from tissue samples (Qiagen Kit) and having an RNA integrity number greater than 7.0 was hybridized to Agilent whole-genome microarrays of rhesus macaque (part G2519F) according to the manufacturer’s instructions.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 November 2011; accepted 11 January 2012.

Published online 29 February 2012.

- Roger, V. L. *et al.* Heart disease and stroke statistics—2011 update: a report from the American Heart Association. *Circulation* **123**, e18–e209 (2011).
- Anon. The bitterest pill. *Nature* **444**, 532–533 (2006).
- O’Collins, V. E. *et al.* 1,026 experimental treatments in acute stroke. *Ann. Neurol.* **59**, 467–477 (2006).
- Sacchetti, M. L. Is it time to definitely abandon neuroprotection in acute ischemic stroke? *Stroke* **39**, 1659–1660 (2008).
- Courtine, G. *et al.* Can experiments in nonhuman primates expedite the translation of treatments for spinal cord injury in humans? *Nature Med.* **13**, 561–566 (2007).
- Enard, D., Depaulis, F. & Roest Crollius, H. Human and non-human primate genomes share hotspots of positive selection. *PLoS Genet.* **6**, e1000840 (2010).
- Aarts, M. *et al.* Treatment of ischemic brain damage by perturbing NMDA receptor-PSD-95 protein interactions. *Science* **298**, 846–850 (2002).
- Sattler, R. *et al.* Specific coupling of NMDA receptor activation to nitric oxide neurotoxicity by PSD-95 protein. *Science* **284**, 1845–1848 (1999).
- Soriano, F. X. *et al.* Specific targeting of pro-death NMDA receptor signals with differing reliance on the NR2B PDZ ligand. *J. Neurosci.* **28**, 10696–10710 (2008).
- Sun, H. S. *et al.* Effectiveness of PSD-95 inhibitors in permanent and transient focal ischemia in the rat. *Stroke* **39**, 2544–2553 (2008).
- Roitberg, B. *et al.* Chronic ischemic stroke model in cynomolgus monkeys: behavioral, neuroimaging and anatomical study. *Neurol. Res.* **25**, 68–78 (2003).
- Fisher, M. The ischemic penumbra: identification, evolution and treatment concepts. *Cerebrovasc. Dis.* **17** (suppl. 1), 1–6 (2004).
- Bardutzky, J. *et al.* Characterizing tissue fate after transient cerebral ischemia of varying duration using quantitative diffusion and perfusion imaging. *Stroke* **38**, 1336–1344 (2007).
- Fisher, M. *et al.* Update of the stroke therapy academic industry roundtable preclinical recommendations. *Stroke* **40**, 2244–2250 (2009).
- Committee, S. Recommendations for standards regarding preclinical neuroprotective and restorative drug development. *Stroke* **30**, 2752–2758 (1999).
- Kornau, H. C., Schenker, L. T., Kennedy, M. B. & Seeburg, P. H. Domain interaction between NMDA receptor subunits and the postsynaptic density protein PSD-95. *Science* **269**, 1737–1740 (1995).
- Cui, H. *et al.* PDZ protein interactions underlying NMDA-receptor-mediated excitotoxicity and neuroprotection by PSD-95 inhibitors. *J. Neurosci.* **27**, 9901–9915 (2007).
- Bratane, B. T. *et al.* Neuroprotection by freezing ischemic penumbra evolution without cerebral blood flow augmentation with a postsynaptic density-95 protein inhibitor. *Stroke* **42**, 3265–3270 (2011).
- Marshall, J. W. & Ridley, R. M. Assessment of cognitive and motor deficits in a marmoset model of stroke. *ILAR J.* **44**, 153–160 (2003).
- Reagan-Shaw, S., Nihal, M. & Ahmad, N. Dose translation from animal to human studies revisited. *FASEB J.* **22**, 659–661 (2008).
- Brott, T. *et al.* Measurements of acute cerebral infarction: a clinical examination scale. *Stroke* **20**, 864–870 (1989).
- Johnston, K. C. *et al.* Validation of an acute ischemic stroke model: does diffusion-weighted imaging lesion volume offer a clinically significant improvement in prediction of outcome? *Stroke* **38**, 1820–1825 (2007).
- Hand, P. J. *et al.* MR diffusion-weighted imaging and outcome prediction after ischemic stroke. *Neurology* **66**, 1159–1163 (2006).
- Yanagihara, T. Experimental stroke in gerbils: effect on translation and transcription. *Brain Res.* **158**, 435–444 (1978).
- Hacke, W. *et al.* Association of outcome with early stroke treatment: pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke trials. *Lancet* **363**, 768–774 (2004).
- Hacke, W. *et al.* Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke. *N. Engl. J. Med.* **359**, 1317–1329 (2008).
- Sanossian, N. *et al.* Simultaneous ring voice-over-Internet phone system enables rapid physician elicitation of explicit informed consent in prehospital stroke treatment trials. *Cerebrovasc. Dis.* **28**, 539–544 (2009).
- The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. *N. Engl. J. Med.* **333**, 1581–1587 (1995).
- Savitz, S. I. & Fisher, M. Future of neuroprotection for acute stroke: in the aftermath of the SAINT trials. *Ann. Neurol.* **61**, 396–402 (2007).
- Savitz, S. I. A critical appraisal of the NXY-059 neuroprotection studies for acute stroke: a need for more rigorous testing of neuroprotective agents in animal models of stroke. *Exp. Neurol.* **205**, 20–25 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants to M.T. from the Canadian Stroke Network and the Heart and Stroke Foundation of Ontario, grant NA 6988. D.J.C. is a recipient of a Canadian Stroke Network postdoctoral research fellowship. M.T. is a Canada Research Chair (Tier 1) in Translational Stroke Research. We thank M. Madden, B. Maloo, A. Goldstein, B. Madeira, W. Foltz and Z. Lu for assistance. We thank M. Salter and M. Hill for a review of the manuscript.

Author Contributions D.C. and M.T. performed the experimental procedures, collected and analysed the data and drafted the manuscript. L.T. performed experimental procedures and data collection.

Author Information Microarray data are deposited in National Center for Biotechnology Information (NCBI) Gene Expression Omnibus under accession number GSE35589. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.T. (mike.tymianski@uhn.ca).

METHODS

Animals. Male captive-bred cynomolgus macaques (2.4–5.0 kg) were pair housed in an environmentally controlled and enclosed primate colony of 10–30 animals on a 12-h light–dark cycle (light on 06:00 to 18:00). Animals were housed in 115 cm × 115 cm × 200 cm home cages and had access to a common recreation space during light hours. The colony was supervised daily by an experienced team consisting of a veterinarian, two veterinary technicians, one laboratory technician and one graduate student. Animals were provided with water *ad libitum*, daily complete diet in the form of Monkey Chow (Purina) and mixed dietary enrichment in the form of nuts, fresh fruit and vegetables throughout the day. Environmental enrichment in the form of puzzles, primate-specific toys and audiovisual media was provided during light hours. Animals were fasted for 12 h before administration of anaesthesia or behavioural testing.

Stroke model. Animals were anaesthetized (isoflurane 1.0–2.5%), intubated and ventilated. Non-invasive monitoring included blood pressure by leg cuff, end-tidal CO₂, O₂ saturation, electrocardiogram and temperature by rectal probe. Temperature was maintained (37 ± 0.5 °C) by heating blanket. A femoral arterial line was used to monitor blood pressure and blood gases. MCAO in cynomolgus macaques (3.0–4.0 kg) was performed using a right pterional craniotomy and occluding the right MCA in the Sylvian fissure³¹ with a 5-mm titanium aneurysm clip either proximal or distal to the orbitofrontal branch and origin of lenticulo-striate arteries, thus producing a stroke with a large penumbra (MCAO^(+P)) or small penumbra (MCAO^(-P)), respectively (main text and Supplementary Fig. 1). The final infarct volume produced in the MCAO^(+P) model was approximately 30% smaller than that produced with MCAO^(-P) owing to the presence of leptomeningeal collaterals. At the end of the MCAO, the aneurysm clip was removed to restore blood flow.

Defining ischaemic penumbra tissue. Penumbra tissue was operationally defined as tissue that was not yet infarcted at the time of tissue collection, but which consistently went on to later infarction. Because the penumbra might be variable in macaques, eight 2 mm × 2 mm biopsies were taken at either 1 or 6 h after MCAO from cortex across the entire MCA vascular distribution ipsilateral to the stroke and contralateral to the stroke from sites mirroring those taken from ischaemic cortex (Supplementary Fig. 4). Biopsied positions were photographed (Supplementary Fig. 4a). To determine which of the eight biopsies represent penumbral tissue, the animals were transferred to a 7 T MRI scanner and DWI, T₂ and perfusion imaging was performed within 15 min after the biopsy and at 5.75 h. Penumbra tissue at 1 h was defined as tissue devoid of infarction that progressed to infarction at 5.75 h by DWI (Supplementary Fig. 4b–d). Penumbral tissue at 6 h was defined as tissue within the confines of the MR perfusion defect but without demonstrable DWI or T₂ hyperintensity. The use of MRI to define the penumbra of NHPs is essential as the amount of salvageable tissue shrinks significantly by 6 h in the MCAO^(-P) model.

General MRI procedures. Imaging was performed on a 7 T Bruker BioSpec system running Paravision 4.0 software and using a B-GA20S gradient coil. A 15.5-cm inner diameter quadrature transmit/receive volume coil was used for NHP scans. NHPs were intubated, ventilated and imaged prone. Physiological monitoring was maintained throughout. The protocol provides stacks of two-dimensional T₂-weighted, perfusion- and diffusion-weighted images in an axial plane. T₂-weighted imaging uses the rapid acquisition with relaxation enhancement (RARE) method, also termed fast-spin echo (echo time/repetition time (T_E/T_R) = 84/5,000 ms, rare factor = 14, 225 × 225 matrix over a 9-cm field of view for 400 μm × 400 μm × 1500 μm resolution). Diffusion-weighted imaging uses a spin-echo multi-shot echo-planar imaging (EPI) technique (T_E/T_R = 32/10,000 ms, nine EPI shots, 250 kHz bandwidth, three orthogonal diffusion directions at b = 1,000 s mm⁻², ten averages with a 180 × 180 matrix over a 9-cm field of view for 500 μm × 500 μm × 1500 μm resolution). Perfusion imaging was performed using a dynamic, contrast-enhanced, susceptibility-weighted perfusion method (T₂*EPI, T_E = 18 ms, two EPI shots, 2-s temporal resolution and 90 repetitions, 180 × 180 matrix over a 9-cm field of view for 700 μm × 700 μm × 1500 μm resolution) over five contiguous slices. For perfusion scans, a gadolinium (0.1 mmol kg⁻¹) bolus was injected intravenously, starting on the third repetition with a total injection time of 7 s through a peripheral intravenous port. Diffusion images were post-processed in MATLAB (Natick) to generate an average image from three b = 1,000 s mm⁻² images and to calculate an apparent diffusion coefficient map. Stroke volumes were calculated using ITK-Snap contouring software (Pittsburgh) with stacks of average diffusion images reconstructed in three dimensions. Perfusion imaging was processed using PerfTool software³² to produce cerebral blood flow maps.

Microarrays. RNA isolated from tissue samples (Qiagen Kit) and having an RNA integrity number greater than 7.0 were hybridized to Agilent whole-genome microarrays of rhesus macaque (part G2519F) according to the manufacturer's instructions. Cynomolgus arrays are not commercially available, but rhesus and cynomolgus macaques are closely related on the phylogenetic tree³³ and the use of NHP- and human-specific microarrays to study related Old World monkeys is customary³⁴. The rhesus genome is sequenced and is integrated into the bioinformatics packages used in our analysis³⁵. Several controls are embedded. First, each array contains

43,803 gene transcripts, of which most genes are duplicated. Additionally, each array includes positive and negative controls for the hybridization reaction and control spots to confirm orientation of the array and order and position of probe sets during scanning. Two-colour arrays were hybridized with experimental samples on one channel and a common Universal RNA Reference sample on the second channel allowing additional corrections for heterogeneity in hybridization and variation in background staining within arrays as well as a means to normalize data between arrays. This provided a control for inter-array comparisons. Intra- and inter-array corrections were performed before generating a ratio of fluorescence of sample: universal control for each complementary DNA expressed sequence tag. Gene lists were generated by comparing differential gene expression (stroke versus contralateral control brain) within each group at 1 and 6 h. Significantly, differentially expressed genes were defined as those with fold change ≥ 2 and P ≤ 0.05. Mean fold change for each differentially regulated expressed sequence tag in each treatment group was included in an unsupervised hierarchical cluster analysis (Agilent GeneSpring GX 10) to provide a graphical display of gene expression data (Fig. 3). Unique and common differentially expressed genes from each treatment group were further reviewed by constructing Venn diagrams from the lists of significantly differentially expressed genes (Venny online tool, bioinfogp.cnb.csic.es/tools/venny/index.html). Significantly differentially expressed genes were also included in a pathway analysis using Pathway-Express, a tool that accounts for gene enrichment for KEGG pathways and ranks pathways based on the number of genes from each pathway present and the relative expression of these genes³⁶.

Experimental design and statistical analysis. The stroke experiments were performed in compliance with the 'recommendations for ensuring good scientific enquiry' of the STAIR committee³⁴. Sample size determinations were based on the desire to detect a 40% difference in infarct volumes between drug and placebo based on the final T₂-weighted MRI at a power of 0.8, alpha = 0.05 and an assumed standard deviation of 30% of group means. Primary analysis was based on an intent-to-treat approach, with no exclusions of any animals enrolled. The cynomolgus macaques were block-randomized to treatment with drug or placebo (vehicle only). The investigators responsible for the induction, maintenance and reversal of ischaemia, for decisions about the care of (including the early killing of) experimental animals and for assessment of all outcomes were blinded to the experimental group to which an animal belonged. Differences between groups were measured using Student's *t*-test, or repeated-measures ANOVA, as required. Missing values due to premature death or inability to complete a task were imputed to reflect the worst score achievable on the task, or the maximum possible stroke volume as defined by largest infarct volume achieved across all animals.

Neurological assessments. These were conducted using the previously validated NHPSS³¹ and a sensorimotor battery of tasks including the hill and valley task, two-tube task and six-well task¹⁹. The NHPSS score is a composite of ratings of state of consciousness, defence reaction, grasp reflex, extremity movement, gait, circling, bradykinesia, balance, neglect, visual field cut/hemianopsia and facial weakness, many of which are incorporated in the National Institutes of Health Stroke Scoring system in humans. From a total of 41 points, 0 corresponds to normal behaviour and 41 to severe bilateral neurological impairment. The remaining tests were modified from assays developed for the common marmoset (*Callithrix jacchus*) as described elsewhere^{19,37}. In addition to evaluating finer sensorimotor functions, they also test extinction and perceptual spatial impairment/neglect¹⁹. In pilot experiments in five macaques subjected to a 90-min MCAO, NHPSS results demonstrated an initial peak in score (mean 36.3, SEM = 5.7) that persisted for the first 36 h and then gradually dropped to a plateau between 14 and 30 days (mean = 14.36, SEM = 3.2). Sensorimotor testing revealed that animals had severe left spatiotemporal neglect and left hemiparesis that showed minor recovery over time at 7 and 30 days after stroke. These deficits were evident as significant delays in completion of six-well (mean delay of 7.8× and 5.33× baseline) and hill and valley tasks for the left arm (mean delays of 8.2× and 6.4× baseline on valley segment and 7.6× and 5.8× baseline on hill segment).

- Findlay, J. M., Macdonald, R. L., Weir, B. K. & Grace, M. G. Surgical manipulation of primate cerebral arteries in established vasospasm. *J. Neurosurg.* **75**, 425–432 (1991).
- Kosior, J. C. & Frayne, R. PerfTool: a software platform for investigating bolus-tracking perfusion imaging quantification strategies. *J. Magn. Reson. Imaging* **25**, 653–659 (2007).
- Stewart, C. B. & Disotell, T. R. Primate evolution – in and out of Africa. *Curr. Biol.* **8**, R582–R588 (1998).
- Kobasa, D. *et al.* Aberrant innate immune response in lethal infection of macaques with the 1918 influenza virus. *Nature* **445**, 319–323 (2007).
- Gibbs, R. A. *et al.* Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222–234 (2007).
- Draghici, S. *et al.* A systems biology approach for pathway level analysis. *Genome Res.* **17**, 1537–1545 (2007).
- Marshall, J. W. & Ridley, R. M. Assessment of functional impairment following permanent middle cerebral artery occlusion in a non-human primate species. *Neurodegeneration* **5**, 275–286 (1996).

Insights into hominid evolution from the gorilla genome sequence

Aylwyn Scally¹, Julien Y. Dutheil^{2†}, LaDeana W. Hillier³, Gregory E. Jordan⁴, Ian Goodhead^{1†}, Javier Herrero⁴, Asger Hobolth², Tuuli Lappalainen⁵, Thomas Mailund², Tomas Marques-Bonet^{3,6,7}, Shane McCarthy¹, Stephen H. Montgomery⁸, Petra C. Schwalie⁴, Y. Amy Tang¹, Michelle C. Ward^{9,10}, Yali Xue¹, Bryndis Yngvadottir^{1†}, Can Alkan^{3,11}, Lars N. Andersen², Qasim Ayub¹, Edward V. Ball¹², Kathryn Beal⁴, Brenda J. Bradley^{8,13}, Yuan Chen¹, Chris M. Clee¹, Stephen Fitzgerald⁴, Tina A. Graves¹⁴, Yong Gu¹, Paul Heath¹, Andreas Heger¹⁵, Emre Karakoc³, Anja Kolb-Kokocinski¹, Gavin K. Laird¹, Gerton Lunter¹⁶, Stephen Meader¹⁵, Matthew Mort¹², James C. Mullikin¹⁷, Kasper Munch², Timothy D. O'Connor⁸, Andrew D. Phillips¹², Javier Prado-Martinez⁶, Anthony S. Rogers^{1†}, Saba Sajjadian³, Dominic Schmidt^{9,10}, Katy Shaw¹², Jared T. Simpson¹², Peter D. Stenson¹², Daniel J. Turner^{1†}, Linda Vigilant¹⁸, Albert J. Vilella⁴, Weldon Whitener¹, Baoli Zhu^{19†}, David N. Cooper¹², Pieter de Jong¹⁹, Emmanouil T. Dermitzakis⁵, Evan E. Eichler^{3,11}, Paul Flicek⁴, Nick Goldman⁴, Nicholas I. Mundy⁸, Zemin Ning¹, Duncan T. Odum^{1,9,10}, Chris P. Ponting¹⁵, Michael A. Quail¹, Oliver A. Ryder²⁰, Stephen M. Searle¹, Wesley C. Warren¹⁴, Richard K. Wilson¹⁴, Mikkel H. Schierup², Jane Rogers^{1†}, Chris Tyler-Smith¹ & Richard Durbin¹

Gorillas are humans' closest living relatives after chimpanzees, and are of comparable importance for the study of human origins and evolution. Here we present the assembly and analysis of a genome sequence for the western lowland gorilla, and compare the whole genomes of all extant great ape genera. We propose a synthesis of genetic and fossil evidence consistent with placing the human–chimpanzee and human–chimpanzee–gorilla speciation events at approximately 6 and 10 million years ago. In 30% of the genome, gorilla is closer to human or chimpanzee than the latter are to each other; this is rarer around coding genes, indicating pervasive selection throughout great ape evolution, and has functional consequences in gene expression. A comparison of protein coding genes reveals approximately 500 genes showing accelerated evolution on each of the gorilla, human and chimpanzee lineages, and evidence for parallel acceleration, particularly of genes involved in hearing. We also compare the western and eastern gorilla species, estimating an average sequence divergence time 1.75 million years ago, but with evidence for more recent genetic exchange and a population bottleneck in the eastern species. The use of the genome sequence in these and future analyses will promote a deeper understanding of great ape biology and evolution.

Humans share many elements of their anatomy and physiology with both gorillas and chimpanzees, and our similarity to these species was emphasized by Darwin and Huxley in the first evolutionary accounts of human origins¹. Molecular studies confirmed that we are closer to the African apes than to orang-utans, and on average closer to chimpanzees than gorillas² (Fig. 1a). Subsequent analyses have explored functional differences between the great apes and their relevance to human evolution, assisted recently by reference genome sequences for chimpanzee³ and orang-utan⁴. Here we provide a reference assembly and initial analysis of the gorilla genome sequence, establishing a foundation for the further study of great ape evolution and genetics.

Recent technological developments have substantially reduced the costs of sequencing, but the assembly of a whole vertebrate genome

remains a challenging computational problem. We generated a reference assembly from a single female western lowland gorilla (*Gorilla gorilla gorilla*) named Kamilah, using 5.4×10^9 base pairs (5.4 Gbp) of capillary sequence combined with 166.8 Gbp of Illumina read pairs (Methods Summary). Genes, transcripts and predictions of gene orthologues and paralogues were annotated by Ensembl⁵, and additional analysis found evidence for 498 functional long (>200-bp) intergenic RNA transcripts. Table 1 summarizes the assembly and annotation properties. An assessment of assembly quality using finished fosmid sequences found that typical (N50; see Table 1 for definition) stretches of error-free sequence are 7.2 kbp in length, with errors tending to be clustered in repetitive regions. Outside repeat masked regions and away from contig ends, the total

¹Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK. ²Bioinformatics Research Center, Aarhus University, C.F. Møllers Allé 8, 8000 Aarhus C, Denmark. ³Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington 98195, USA. ⁴European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SD, UK. ⁵Department of Genetic Medicine and Development, University of Geneva Medical School, Rue Michel-Servet 1, 1211 Geneva 4, Switzerland. ⁶Institut de Biologia Evolutiva (UPF-CSIC), 08003 Barcelona, Catalonia, Spain. ⁷Institució Catalana de Recerca i Estudis Avançats, ICREA, 08010 Barcelona, Spain. ⁸Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK. ⁹University of Cambridge, Department of Oncology, Hutchison/MRC Research Centre, Hills Road, Cambridge CB2 0XZ, UK. ¹⁰Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. ¹¹Howard Hughes Medical Institute, University of Washington, Seattle, Washington 20815-6789, USA. ¹²Institute of Medical Genetics, Cardiff University, Heath Park, Cardiff CF14 4XN, UK. ¹³Department of Anthropology, Yale University, 10 Sachem Street, New Haven, Connecticut 06511, USA. ¹⁴The Genome Institute at Washington University, Washington University School of Medicine, Saint Louis, Missouri 63108, USA. ¹⁵MRC Functional Genomics Unit, University of Oxford, Department of Physiology, Anatomy and Genetics, South Parks Road, Oxford OX1 3QX, UK. ¹⁶Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK. ¹⁷Comparative Genomics Unit, Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, 20892-2152, USA. ¹⁸Max Planck Institute for Evolutionary Anthropology, Primate Department, Deutscher Platz 6, Leipzig 04103, Germany. ¹⁹Children's Hospital Oakland Research Institute, Oakland, California 94609, USA. ²⁰San Diego Zoo's Institute for Conservation Research, Escondido, California 92027, USA. [†]Present addresses: Institut des Sciences de l'Évolution – Montpellier (I.S.E.-M.), Université de Montpellier II – CC 064, 34095 Montpellier Cedex 05, France (J.Y.D.); Centre for Genomic Research, Institute of Integrative Biology, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK (I.G.); Division of Biological Anthropology, University of Cambridge, Fitzwilliam Street, Cambridge CB2 1QH, UK (B.Y.); EASIH, University of Cambridge, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK (A.S.R.); Oxford Nanopore Technologies, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford OX4 4GA, UK (D.J.T.); Institute of Microbiology, Chinese Academy of Sciences, Datun Road, Chaoyang District, Beijing 100101, China (B.Z.); The Genome Analysis Centre, Norwich Research Park, Norwich NR4 7UH, UK (J.R.).

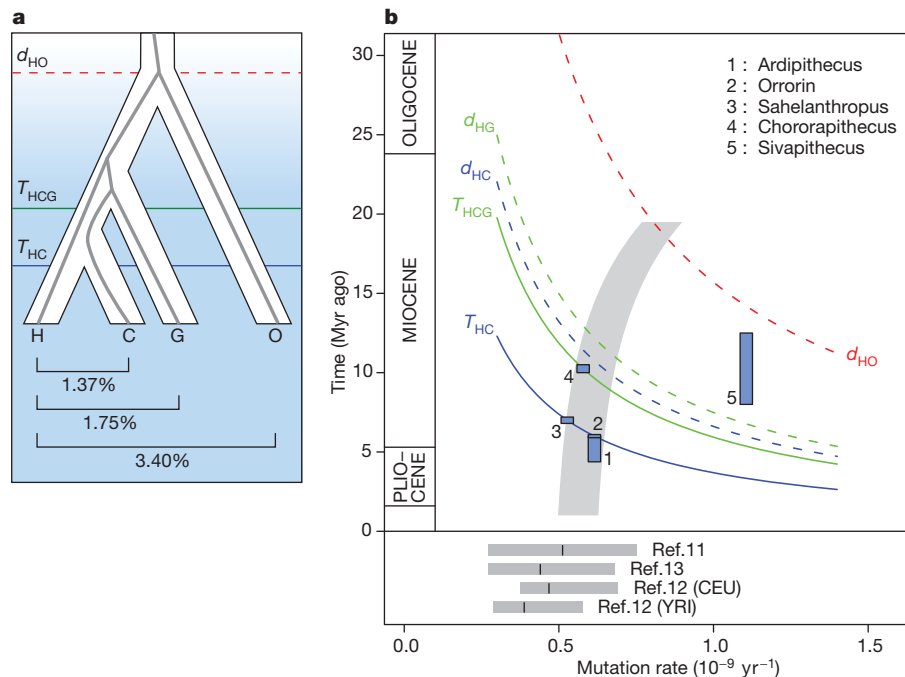


Figure 1 | Speciation of the great apes. **a**, Phylogeny of the great ape family, showing the speciation of human (H), chimpanzee (C), gorilla (G) and orang-utan (O). Horizontal lines indicate speciation times within the hominine subfamily and the sequence divergence time between human and orang-utan. Interior grey lines illustrate an example of incomplete lineage sorting at a particular genetic locus—in this case (((C, G), H), O) rather than (((H, C), G), O). Below are mean nucleotide divergences between human and the other great apes from the EPO alignment. **b**, Great ape speciation and divergence times. Upper panel, solid lines show how times for the HC and HCG speciation events estimated by CoalHMM vary with average mutation rate; dashed lines show the corresponding average sequence divergence times, as well as the HO sequence divergence. Blue blocks represent

rate of single-base and indel errors is 0.13 per kbp. See Supplementary Information for further details.

We also collected less extensive sequence data for three other gorillas, to enable a comparison of species within the *Gorilla* genus. Gorillas survive today only within several isolated and endangered populations whose evolutionary relationships are uncertain. In addition to Kamilah, our analysis included two western lowland gorillas, Kwanza (male) and EB(JC) (female), and one eastern lowland gorilla, Mukisi (male).

Speciation of the great apes

We included the Kamilah assembly with human, chimpanzee (*Pan troglodytes*), orang-utan (*Pongo abelii*) and macaque (*Macaca mulatta*) in a five-way whole-genome alignment using the Ensembl EPO pipeline⁶ (Supplementary Table 3.2). Filtering out low-quality regions of the chimpanzee assembly and regions with many alignment gaps, we obtained 2.01 Gbp of 1:1:1:1 great ape orthologous alignment blocks, to which we then applied a coalescent inference model,

Table 1 | Assembly and annotation statistics

Assembly	Annotation
Total length	3,041,976,159 bp
Contigs	465,847
Total contig length	2,829,670,843 bp
Placed contig length	2,712,844,129 bp
Unplaced contig length	116,826,714 bp
Max. contig length	191,556 bp
Contig N50	11.8 kbp
Scaffolds	22,164
Max. scaffold length	10,247,101 bp
Scaffold N50	914 kbp
Protein-coding genes	20,962
Pseudogenes	1,553
RNA genes	6,701
Gene exons	237,216
Gene transcripts	35,727
lincRNA transcripts	498

N50: 50% of the genome is in fragments of this length or longer; lincRNA: long intergenic non-coding RNA.

hominid fossil species (key at top right): each has a vertical extent spanning the range of dates estimated for it in the literature^{9,50}, and a horizontal position at the maximum mutation rate consistent both with its proposed phylogenetic position and the CoalHMM estimates (including some allowance for ancestral polymorphism in the case of *Sivapithecus*). The grey shaded region shows that an increase in mutation rate going back in time can accommodate present-day estimates, fossil hypotheses, and a middle Miocene speciation for orang-utan. Lower panel, estimates of the average mutation rate in present-day humans^{11–13}; grey bars show 95% confidence intervals, with black lines at the means. Estimates were made by the 1000 Genomes Project for trios of European (CEU) and Yoruban African (YRI) ancestry.

CoalHMM, to estimate the timescales and population sizes involved in the speciation of the hominines (African great apes; see Supplementary Table 1.1 for terminology), with orang-utan as an out-group (Supplementary Information).

Two issues need to be addressed in interpreting the results from CoalHMM (Supplementary Table 4.2). First, the results themselves are obtained in units of sequence divergence rather than years, and so need to be scaled by an appropriate yearly mutation rate. Second, as with any model, CoalHMM makes several simplifying assumptions whose consequences we need to understand in the context of realistic demography. We discuss these issues in turn.

Using a rate of 10^{-9} mutations per bp per year, derived from fossil calibration of the human-macaque sequence divergence and as used in previous calculations, CoalHMM's results would correspond to speciation time estimates T_{HC} (for human-chimpanzee) and T_{HCG} (for human-chimpanzee-gorilla) of 3.7 and 5.95 Myr ago, respectively (Fig. 1b). These dates are consistent with other recent molecular estimates^{7,8}, but are at variance with certain aspects of the fossil record, including several fossils which have been proposed—though not universally accepted⁹—to be hominins, and therefore to postdate the human-chimpanzee split (Fig. 1b). Indeed, the relationship between molecular and fossil evidence has remained difficult to resolve despite the accumulation of genetic data¹⁰. Direct estimates of the per-generation mutation rate in modern human populations, based on the incidence of disease-causing mutations¹¹ or sequencing of familial trios^{12,13}, indicate that a lower value of $(0.5–0.6) \times 10^{-9}$ bp⁻¹ yr⁻¹ is plausible (based on average hominine generation times of 20–25 yr). This would give substantially older estimates of approximately 6 and 10 Myr ago for T_{HC} and T_{HCG} , potentially in better agreement with the fossil record.

However, this timetable for hominine speciation must also be reconciled with older events, such as the speciation of orang-utan, which is thought to have occurred no earlier than the Middle Miocene (12–16 Myr ago), as fossil apes before that differ substantially from what we might expect of an early great ape¹⁴. This is possible if we allow for mutation rates changing over time, with a mutation rate of around $1 \times 10^{-9} \text{ bp}^{-1} \text{ yr}^{-1}$ in the common ancestor of great apes, decreasing to lower values in all extant species (Fig. 1b). Comparable changes in mutation rate have been observed previously in primate evolution on larger timescales, including an approximately 30% branch length decrease in humans compared to baboons since their common ancestor¹⁵. A decrease within the great apes is also a predicted consequence of the observed increase in body sizes over this time period and the association of small size with shorter generation times in other primates¹⁶, and is consistent with deviations from a molecular clock seen in sequence divergences of the great apes and macaque (Supplementary Table 3.3). We discuss these and other constraints on estimates of great ape speciation times in the Supplementary Information. However we note that *Sahelanthropus* and *Chororapithecus* remain difficult to incorporate in this model, and can be accommodated as hominin and gorilla genera only if most of the decrease occurred early in great ape evolution.

An alternative explanation for the apparent discrepancy in fossil and genetic dates (leaving aside the issue of whether fossil taxa have been correctly placed) is that ancestral demography may have affected the genetic inferences. Certainly CoalHMM's model does not fit the data in all respects. Perhaps most importantly, it assumes that ancestral population sizes are constant in time and that no gene flow occurred between separated populations, approximations that may not hold in reality. Simulations (details in Supplementary Information) suggest that an ancestral population bottleneck would have had limited impact on the inference of T_{HC} , its influence being captured largely by changes in the model's effective population size. Under conditions of genetic exchange between populations after the main separation of the chimpanzee and human lineages, the speciation time estimated by CoalHMM represents an average weighted by gene flow over the period of separation. This means in some cases it can be substantially older than the date of most recent exchange. However it would only be more recent than the speciation time inferred from fossils if there had been strong gene flow between populations after the development of derived fossil characteristics. To the extent that this is plausible, for example as part of a non-allopatric speciation process, it constitutes an alternative explanation for the dating discrepancy without requiring a change in mutation rate.

In summary, although whole-genome comparisons can be strongly conclusive about the ordering of speciation events, the inability to observe past mutation rates means that the timing of events from genetic data remains uncertain. In our view, possible variation in mutation rates allows hominid genomic data to be consistent with values of T_{HC} from 5.5 to 7 Myr ago and T_{HCG} from 8.5 to 12 Myr ago, with ancestral demographic structure potentially adding inherent ambiguity to both events. Better resolution may come from further integrated analysis of fossil and genetic evidence.

Incomplete lineage sorting and selection

The genealogy relating human (H), chimpanzee (C) and gorilla (G) varies between loci across the genome. CoalHMM explicitly models this and infers the genealogy at each position: either the standard ((H,C),G) relationship or the alternatives ((H,G),C) or ((C,G),H), which are the consequences of incomplete lineage sorting (ILS) in the ancestral human–chimpanzee population. We can use the pattern of ILS to explore evolutionary forces during the human–chimpanzee–gorilla speciation period. Across the genome we find 30% of bases exhibiting ILS, with no significant difference between the number sorting as ((H,G),C) and ((C,G),H). However, the fraction of ILS varies with respect to genomic position (Fig. 2a) by more than

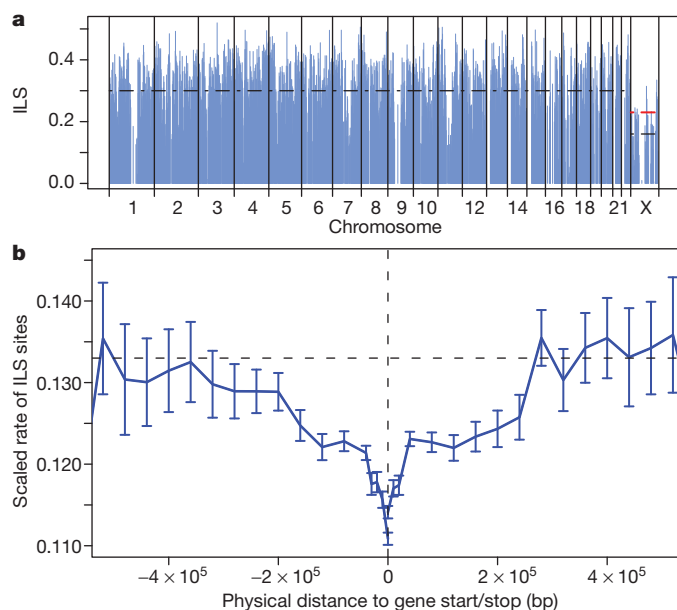


Figure 2 | Genome-wide incomplete lineage sorting (ILS) and selection. **a**, Variation in ILS. Each vertical blue line represents the fraction of ILS between human, chimpanzee and gorilla estimated in a 1-Mbp region. Dashed black lines show the average ILS across the autosomes and on X; the red line shows the expected ILS on X, given the autosomal average and assuming neutral evolution. **b**, Reduction in ILS around protein coding genes. The blue line shows the mean rate of ILS sites normalized by mutation rate as a function of distance upstream or downstream of the nearest gene (see Supplementary Information). The horizontal dashed line indicates the average value outside 300 kbp from the nearest gene; error bars are s.e.m.

expected under a model of genome-wide neutral evolution (Supplementary Fig. 5.1). This variation reflects local differences in the ancestral effective population size N_e during the period between the gorilla and chimpanzee speciation events, most probably due to natural selection reducing N_e and making ILS less likely. Within coding exons mean ILS drops to 22%, and the suppression of ILS extends out to several hundred kbp from coding genes, evident even in raw site patterns before any model inference (Fig. 2b). An analysis of ILS sites in human segmental duplications suggests that assembly errors do not contribute significantly to this signal (Supplementary Information). We therefore attribute it to the effects of linkage around selected mutations, most probably in the form of background selection¹⁷, observing that it is greater around genes with lower ratios of non-synonymous to synonymous mutation rates (dN/dS) (Supplementary Fig. 8.4). Given that more than 90% of the genome lies within 300 kbp of a coding gene, and noting the similar phenomenon reported for recent human evolution¹², this supports the suggestion that selection has affected almost all of the genome throughout hominid evolution¹⁸.

In fitting the transitions between genealogies along the alignment, CoalHMM also estimates a regional recombination rate. This is primarily sensitive to ancestral crossover events before human–chimpanzee speciation, yet despite the expectation of rapid turnover in recombination hotspots¹⁹, averaged over 1-Mbp windows there is a good correlation with estimates from present-day crossovers in humans ($R = 0.49$; $P < 10^{-13}$; Supplementary Fig. 5.5), consistent with the conservation of recombination rates between humans and chimpanzees on the 1-Mbp scale¹⁹.

As expected, we see reduced ILS (Fig. 2a) and human–chimpanzee sequence divergence d_{HC} (Supplementary Fig. 6.1) on the X chromosome, corresponding to a difference in N_e between X and the autosomes within the ancestral human–chimpanzee population. Several factors can contribute to this difference²⁰, notably the X chromosome's haploidy in males, which reduces N_e on X by 0.75, enhances purifying selection in males, and reduces the recombination rate, thereby increasing the

effect of selection via linkage. However, sequence divergence is additionally affected by the mutation rate, which is higher in males than in females, further reducing the relative divergence observed on X²¹. Incorporating the ancestral N_e estimates from CoalHMM, we estimate a ratio of 0.87 ± 0.09 between average mutation rates on X and the autosomes on the human–chimpanzee lineage, corresponding to a male/female mutation rate bias $\alpha = 2.3 \pm 0.4$ (details in Supplementary Information). Previous estimates of α in hominids have ranged from 2 to 7 (refs 22, 23). It is possible that some of the higher values, having been estimated from sequence divergence only and in smaller data sets, were inflated by underestimating the suppression of ancestral N_e on X, in particular due to purifying selection.

Our calculation of α assumes that a single speciation time applies across the genome, attributing differences between the X chromosome and autosomes to the factors mentioned above. An alternative model has been proposed²⁴, involving complex speciation, with more recent human–chimpanzee ancestry on X than elsewhere. Given potential confounding factors in demography, selection, mutation rate bias and admixture, our analyses do not discriminate between these models; however if the effective human–chimpanzee separation time on X is indeed reduced in this way it would imply a still lower value of α .

Functional sequence evolution

We looked for loss or gain of unique autosomal sequence within humans, chimpanzees and gorillas by comparing raw sequence data for each in the context of their reference assemblies (Supplementary Information). The total amount is small: 3–7 Mbp per species, distributed genome-wide in fragments no more than a few kbp in length (Supplementary Table 7.1). The vast majority (97%) of such material was also found either in orang-utan or a more distant primate, indicating loss, and consistent with the expectation that gain is driven primarily by duplication (which our analysis excludes). Some fragments found only in one species overlap coding exons in annotated genes: 6 genes in human, 5 in chimpanzee and 9 in gorilla (Supplementary Tables 7.2, 7.3, 7.4), the majority being associated with olfactory receptor proteins or other rapidly evolving functions, such as male fertility and immune response.

We did not assemble a gorilla Y chromosome, but by mapping $\sim 6\times$ reads from the male gorillas Kwanza and Mukisi to the human Y, we identified several regions in which human single-copy material is missing in gorilla, comprising almost 10% of the accessible male-specific region. Across the Y chromosome there is considerable variation in the copy number of shared material, and the pattern of coverage is quite different from that of reads from a male bonobo mapped in the same way (Supplementary Fig. 7.1). Some missing or depleted material overlaps coding genes (Supplementary Table 7.5), including for example *VCY*, a gene expressed specifically in male germ cells which has two copies in human and chimpanzee but apparently only one in gorilla (Supplementary Information). The resulting picture is consistent with rapid structural evolution of the Y chromosome in the great apes, as previously seen in the chimpanzee–human comparison²⁵.

Protein evolution

The Ensembl EPO primate alignment was filtered to produce a high-quality genome-wide set of 11,538 alignments representing orthologous primate coding sequences, which were then scored with codon-based evolutionary models for likelihoods of acceleration or deceleration of dN/dS in the terminal lineages, ancestral branch, and entire hominine subfamily (Supplementary Information). We find that genes with accelerated rates of evolution across hominines are enriched for functions associated with sensory perception, particularly in relation to hearing and brain development (Supplementary Table 8.4g, h). For example, among the most strongly accelerated genes are *OTOF* ($P = 0.0056$), *LOXHD1* ($P < 0.01$) and *GPR98* ($P = 0.0056$), which are all associated with diseases causing human deafness (Supplementary

Table 8.5). *GPR98*, which also shows significant evidence of positive selection under the branch-site test ($P = 0.0081$), is highly expressed in the developing central nervous system. The gene with the strongest evidence for acceleration along the branch leading to hominines is *RNF213* (branch-site $P < 2.9 \times 10^{-9}$), a gene associated with Moyamoya disease in which blood flow to the brain is restricted due to arterial stenosis²⁶. Given that oxygen and glucose consumption scales with total neuron number²⁷, *RNF213* may have played a role in facilitating the evolution of larger brains. Together, these observations are consistent with a major role for adaptive modifications in brain development and sensory perception in hominine evolution.

Turning to lineage-specific selection pressures, we find relatively similar numbers of accelerated genes in humans, chimpanzees and gorillas (663, 562 and 535 respectively at nominal $P < 0.05$, Supplementary Table 8.3a) and genome-wide dN/dS ratios (0.256, 0.249 and 0.239 in purifying sites, Supplementary Table 8.6). These numbers, which reflect variation in historical effective population sizes as well as environmental pressures, reveal a largely uniform landscape of recent hominine gene evolution—in accordance with previously published analyses in human and chimpanzee^{3,28} (Supplementary Table 8.7).

Genes with accelerated rates of evolution along the gorilla lineage are most enriched for a number of developmental terms, including ear, hair follicle, gonad and brain development, and sensory perception of sound. Among the most significantly accelerated genes in gorilla is *EVPL* ($P < 2.2 \times 10^{-5}$), which encodes a component of the cornified envelope of keratinocytes, and may be related to

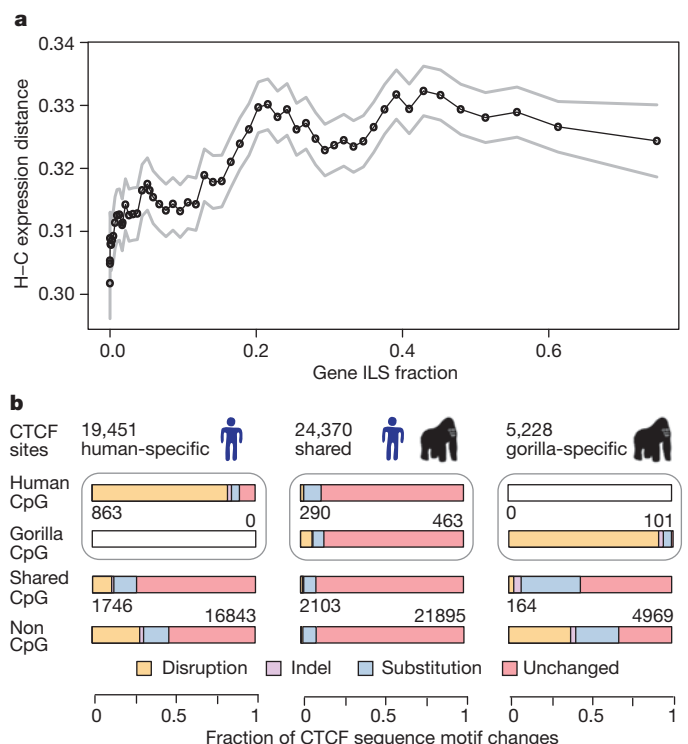


Figure 3 | Differences in expression and regulation. a, Mean gene expression distance between human and chimpanzee as a function of the proportion of ILS sites per gene. Each point represents a sliding window of 900 genes (over genes ordered by ILS fraction); s.d. error limits are shown in grey. **b**, Top row, classification of CTCF sites in the gorilla (EB(JC)) and human (GM12878) LCLs on the basis of species-uniqueness; numbers of alignable CTCF binding sites are shown for each category. Bottom three rows, sequence changes of CTCF motifs embedded in human-specific, shared and gorilla-specific CTCF binding sites located within shared CpG islands, species-specific CpG islands or outside CpG islands. Numbers of CTCF binding sites are shown for each CpG island category. Gorilla and human motif sequences are compared and represented as indels, disruptions (>4-bp gaps) and substitutions.

increased cornification of knuckle pads in gorilla²⁹. Interestingly, gorilla and human both yielded brain-associated terms enriched for accelerated genes, but chimpanzee did not (Supplementary Table 8.4a–c). Genes expressed in the brain or involved in its development have not typically been associated with positive selection in primates, but our results show that multiple great ape lineages show elevated dN/dS in brain-related genes when evaluated against a primate background.

We also identified cases of pairwise parallel evolution among hominines. Human and chimpanzee show the largest amount, with significantly more shared accelerations than expected by chance, whereas gorilla shares more parallel acceleration with human than with chimpanzee across a range of significance thresholds (Supplementary Fig. 8.3). Genes involving hearing are enriched in parallel accelerations for all three pairs, but most strongly in gorilla–human (Supplementary Table 8.4d–f), calling into question a previous link made between accelerated evolution of auditory genes in humans and language evolution²⁸. It is also interesting to note that ear morphology is one of the few external traits in which humans are more similar to gorillas than to chimpanzees³⁰.

Next we considered gene loss and gain. We found 84 cases of gene loss in gorilla due to the acquisition of a premature stop codon, requiring there to be no close paralogue (Supplementary Table 8.8): one such gene is *TEX14*, which codes for an intercellular bridge protein essential for spermatogenesis in mice. Genome-wide analysis of gene gain is confounded by the difficulty in assembling closely related paralogues. We therefore resequenced, by finishing overlapping fosmids, three gene clusters known to be under rapid adaptive evolution in primates: the growth hormone cluster³¹, the PRM clusters involved in sperm function and the APOBEC cluster implicated in molecular adaptation to viral defence. In the growth hormone cluster, we observed four chorionic somatomammotropin (CSH) genes in gorilla compared to three in humans and chimpanzees, with a novel highly similar pair of CSH-like genes in gorilla that share a 3' end similar to human growth hormone *GH2*, suggesting a complex evolutionary history as in other primates³¹. We saw sequence but not gene copy number changes in the PRM and APOBEC clusters (Supplementary Information).

In several cases, a protein variant thought to cause inherited disease in humans³² is the only version found in all three gorillas for which we have genome-wide sequence data (Supplementary Table 8.9). Striking examples are the dementia-associated variant Arg432Cys in the growth factor PGRN and the hypertrophic cardiomyopathy-associated variant Arg153His in the muscle Z disk protein TCAP,

both of which were corroborated by additional capillary sequencing (Supplementary Table 8.10). Why variants that appear to cause disease in humans might be associated with a normal phenotype in gorillas is unknown; possible explanations are compensatory molecular changes elsewhere, or differing environmental conditions. Such variants have also been found in both the chimpanzee and macaque genomes^{3,33}.

Gene transcription and regulation

We carried out an analysis of hominine transcriptome variation using total RNA extracted and sequenced from lymphoblastoid cell lines (LCLs) of one gorilla, two chimpanzees and two bonobos (Supplementary Information), and published RNA sequence data for eight human individuals³⁴. After quantifying reads mapping to exons and genes in each species, we calculated the degree of species-specific expression and splicing in 9,746 1:1:1 expressed orthologous genes. On average, expression levels in human and chimpanzee were more similar to each other than either was to gorilla (Supplementary Fig. 10.2). However this effect is reduced in genes with a higher proportion of ILS sites, which tend to show greater expression distance between humans and chimpanzees (Fig. 3a). More generally, patterns seen in the relative expression distances between the three species showed a significant overlap with those derived from genomic lineage sorting ($P = 0.026$; Supplementary Table 10.4), demonstrating that ILS can be reflected in functional differences between primate species.

We also explored species specific variation in splicing³⁵ by calculating the variance in differential expression of orthologous exons within each gene. In total we found 7% of genes whose between-species variance is significant at the 1% level (based on the distribution of within-human variances, Supplementary Fig. 10.5). For example, Supplementary Fig. 10.6 illustrates gorilla-specific splicing in the *SQLE* gene, involved in steroid metabolism.

We further investigated great ape regulatory evolution by comparing the binding in human and gorilla of CTCF, a protein essential to vertebrate development that is involved in transcriptional regulation, chromatin loop formation and protein scaffolding³⁶. We performed ChIP-seq (chromatin immunoprecipitation sequencing) of CTCF in a gorilla LCL (from EB(JC)), and compared this with matched human experiments³⁷, using the EPO alignments to identify species-specific and shared binding regions (Fig. 3b and Supplementary Information). Consistent with previous results reporting strong CTCF binding conservation³⁸, and in contrast to the rapid turnover of some other transcription factor binding sites³⁹, we found that approximately 70% of

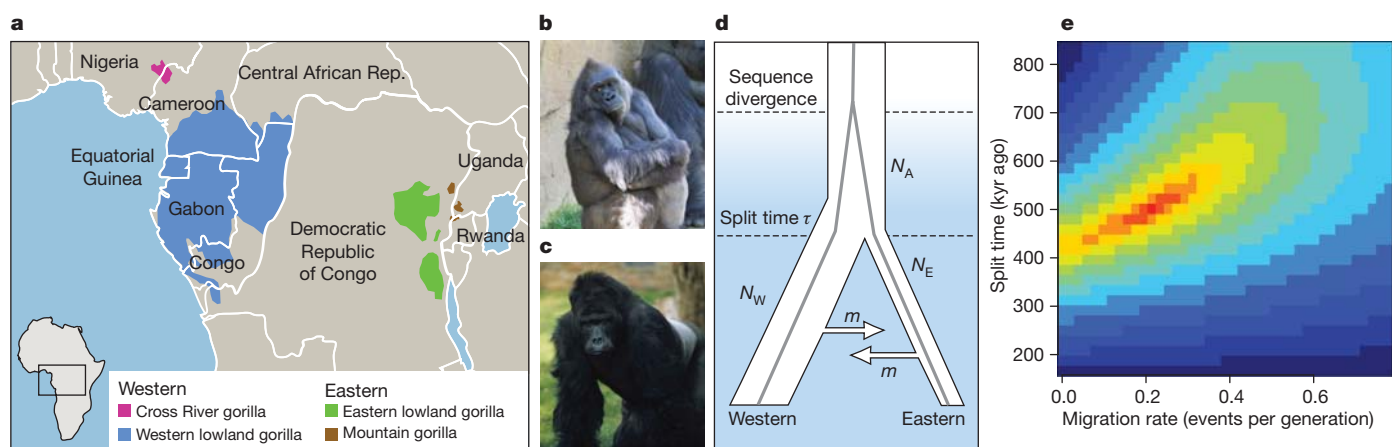


Figure 4 | Gorilla species distribution and divergence. **a**, Distribution of gorilla species in Africa. The western species (*Gorilla gorilla*) comprises two subspecies: western lowland gorillas (*G. gorilla gorilla*) and Cross River gorillas (*G. gorilla diehli*). Similarly, the eastern species (*Gorilla beringei*) is subclassified into eastern lowland gorillas (*G. beringei graueri*) and mountain gorillas (*G. beringei beringei*). (Based on data in ref. 43.) Areas of water are shown pale blue. Inset, area of main map. **b**, Western lowland gorilla Kamilah, source of the

reference assembly (photograph by J.R.). **c**, Eastern lowland gorilla Mukisi (photograph by M. Seres). **d**, Isolation–migration model of the western and eastern species. N_A , N_W and N_E are ancestral, western and eastern effective population sizes; m is the migration rate. **e**, Likelihood surface for migration and split time parameters in the isolation–migration model; colours from blue (minimum) to red (maximum) indicate the magnitude of likelihood.

Table 2 | Nucleotide polymorphism in western and eastern gorillas

Gorilla	Species	Heterozygous site rate (%)	Homozygous site rate (%)	Homozygous:heterozygous ratio
Kamilah	Western lowland	0.189	0.0015	0.008
EB(JC)	Western lowland	0.178	0.10	0.56
Mukisi	Eastern lowland	0.076	0.19	2.5

Rates (columns 3 and 4) are based on variants detected by mapping sequence data to the gorilla reference and filtering sites by depth and mapping quality (Supplementary Information). The homozygosity rate for Kamilah is low (and is effectively an error rate) because the sequence of this gorilla was used for assembly. Reduced heterozygosity in Mukisi is not due to familial inbreeding, as there are no long homozygous stretches.

gorilla CTCF binding regions are shared with human. This compares with around 80% pairwise overlaps between three human LCLs (Supplementary Fig. 11.1a). Binding regions that are shared among all three human individuals are three times more likely to be shared with gorilla than individual-specific regions (Supplementary Fig. 11.1b).

The genomic changes leading to loss of CTCF binding differ between regions within CpG islands and those in the rest of the genome. Losses of CTCF binding outside CpG islands and within species-specific CpG regions co-occur with sequence changes in the binding motif, but for shared CpG islands most binding losses have no corresponding motif sequence change (Fig. 3b). It is possible that DNA methylation differences are driving this effect, as CTCF binding can be abolished by methylation of specific target regions³⁶. Alternatively, CTCF binding within CpG islands may also depend more on other regulators' binding and less on the CTCF motif itself.

Genetic diversity within *Gorilla*

Recent studies of molecular and morphological diversity within the *Gorilla* genus have supported a classification into two species, eastern (*Gorilla beringei*) and western (*Gorilla gorilla*)⁴⁰, with both species further divided into subspecies (Fig. 4a). Although separated today by over 1,000 km, it has been suggested that gene flow has occurred between the eastern and western species since divergence⁴¹. To investigate this, we collected reduced representation sequence data (Supplementary Information) for another female western lowland gorilla, EB(JC), and a male eastern lowland gorilla, Mukisi.

Table 2 summarizes the sequence diversity in these individuals and in Kamilah, based on alignment of sequence data to the gorilla assembly. The ratio of homozygous to heterozygous variant rates for EB(JC) (close to 0.5) is consistent with this gorilla coming from the same population as Kamilah (Supplementary Information), and the rate of heterozygosity for this gorilla matches Kamilah's. Mukisi, on the other hand, has twice the rate of homozygous differences from the assembly, consistent with this gorilla coming from a separate population. Furthermore, heterozygosity in Mukisi is much lower, suggesting a reduced population size in the eastern species. This agrees with previous studies based on fewer loci⁴¹, and also with estimates of present-day numbers in the wild, which indicate that whereas the western lowland subspecies may number up to 200,000 individuals, the eastern population as a whole is around ten times smaller^{42,43}. Because it manifests in genetic diversity, this disparity must have existed for many millennia, and cannot have resulted solely from the current pressure of human activity in central Africa or recent outbreaks of the Ebola virus.

On the basis of an alignment of the EB(JC) and Mukisi data to the human reference sequence and comparing high confidence genotype calls for the two individuals, we estimate a mean sequence divergence time between them of 1.75 Myr ago. However the pattern of shared heterozygosity is not consistent with a clean split between western and eastern gorillas (Supplementary Information). Under a model which allows symmetric genetic exchange between the populations after an initial split (Fig. 4d; Supplementary Information), the maximum likelihood species split time is ~0.5 Myr ago with moderate subsequent exchange of ~0.2 individuals per generation each way between breeding pools, totalling ~5,000 in each direction over 0.5 Myr (Fig. 4e). Different model assumptions and parameterisations would lead to different values. More extensive sampling and sequencing of both gorilla populations will afford better resolution of this issue.

We also collected whole-genome sequence data from an additional male western lowland gorilla (Kwanza) at 12×, and further whole-genome sequence data for (eastern) Mukisi at 7× (Supplementary Information). Differences between the western gorillas and Mukisi represent a combination of inter-individual and inter-species variants. These include 1,615 non-synonymous SNPs (single nucleotide polymorphisms) in 1,326 genes, seven of which have more than four amino acid differences each (Supplementary Table 12.2), among which are two olfactory receptor genes and *EMR3*, implicated in immune and inflammatory responses⁴⁴. Nineteen of the genes annotated in Kamilah carry an apparently homozygous premature stop codon in Mukisi. These include the gene encoding the seminal fluid protein *SEMG2*, implicated in sperm competition and known to be inactivated in some gorillas, where sperm competition is rare⁴⁵. Both *EMR3* and *SEMG2* were corroborated by additional sequencing (Supplementary Tables 12.3, 12.4).

Finally, we investigated genomic duplication in gorilla using a whole-genome shotgun sequence detection method applied to data from the western lowland gorillas Kamilah and Kwanza (Supplementary Information). This revealed a level of private segmental duplication (0.9 Mbp and 1.5 Mbp in the two gorillas) well outside the range found in pairwise comparisons of humans (Supplementary Fig. 13.1), where a value of ~100 kbp is typical between any two individuals⁴⁶. These results suggest greater copy number diversity in gorillas than in humans, consistent with previous observations in the great apes⁴⁷.

Discussion

Since the middle Miocene—an epoch of abundance and diversity for apes throughout Eurasia and Africa—the prevailing pattern of ape evolution has been one of fragmentation and extinction⁴⁸. The present-day distribution of non-human great apes, existing only as endangered and subdivided populations in equatorial forest refugia⁴³, is a legacy of that process. Even humans, now spread around the world and occupying habitats previously inaccessible to any primate, bear the genetic legacy of past population crises. All other branches of the genus *Homo* have passed into extinction. It may be that in the condition of *Gorilla*, *Pan* and *Pongo* we see some echo of our own ancestors before the last 100,000 years, and perhaps a condition experienced many times over several million years of evolution. It is notable that species within at least three of these genera continued to exchange genetic material long after separation^{4,49}, a disposition that may have aided their survival in the face of diminishing numbers. As well as teaching us about human evolution, the study of the great apes connects us to a time when our existence was more tenuous, and in doing so, highlights the importance of protecting and conserving these remarkable species.

METHODS SUMMARY

Assembly. We constructed a hybrid *de novo* assembly combining 5.4 Gbp of capillary read pairs with the contigs from an initial short read assembly of 166.8 Gbp of Illumina paired reads. Improvements in long-range structure were then guided by human homology, placing contigs into scaffolds wherever read pairs confirmed collinearity between gorilla and human. Base-pair contiguity was improved by local reassembly within each scaffold, merging or extending contigs using Illumina read pairs. Finally we used additional Kamilah bacterial artificial chromosome (BAC) and fosmid end pair capillary sequences to provide longer range scaffolding. Base errors were corrected by mapping all Illumina reads back to the assembly and rectifying apparent homozygous variants, while recording the location of heterozygous sites. Further details and other methods are described in Supplementary Information.

Received 16 June 2011; accepted 10 January 2012.

1. Huxley, T. H. *Evidence as to Man's Place in Nature* (Williams & Norgate, 1863).
2. King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
3. Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
4. Locke, D. P. *et al.* Comparative and demographic analysis of orang-utan genomes. *Nature* **469**, 529–533 (2011).
5. Hubbard, T. J. *et al.* Ensembl 2009. *Nucleic Acids Res.* **37**, D690–D697 (2009).
6. Paten, B., Herrero, J., Beal, K., Fitzgerald, S. & Birney, E. Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment with paralogs. *Genome Res.* **18**, 1814–1828 (2008).
7. Bradley, B. J. Reconstructing phylogenies and phenotypes: a molecular view of human evolution. *J. Anat.* **212**, 337–353 (2008).
8. Burgess, R. & Yang, Z. Estimation of hominoid ancestral population sizes under bayesian coalescent models incorporating mutation rate variation and sequencing errors. *Mol. Biol. Evol.* **25**, 1979–1994 (2008).
9. Wood, B. & Harrison, T. The evolutionary context of the first hominins. *Nature* **470**, 347–352 (2011).
10. Steiper, M. E. & Young, N. M. Timing primate evolution: lessons from the discordance between molecular and paleontological estimates. *Evol. Anthropol.* **17**, 179–188 (2008).
11. Lynch, M. Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl Acad. Sci. USA* **107**, 961–968 (2010).
12. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010); correction **473**, 544 (2011).
13. Roach, J. C. *et al.* Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* **328**, 636–639 (2010).
14. Hartwig, W. C. *et al.* *The Primate Fossil Record* (Cambridge Univ. Press, 2002).
15. Kim, S. H., Elango, N., Warden, C., Vigoda, E. & Yi, S. V. Heterogeneous genomic molecular clocks in primates. *PLoS Genet.* **2**, e163 (2006).
16. Fleagle, J. G. *Primate Adaptation and Evolution* 2nd edn (Academic Press, 1998).
17. Charlesworth, D., Morgan, M. T. & Charlesworth, B. Mutation accumulation in finite populations. *J. Hered.* **84**, 321–325 (1993).
18. McVicker, G., Gordon, D., Davis, C. & Green, P. Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet.* **5**, e1000471 (2009).
19. Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**, 321–324 (2005).
20. Vicoso, B. & Charlesworth, B. Evolution on the X chromosome: unusual patterns and processes. *Nature Rev. Genet.* **7**, 645–653 (2006).
21. Ellegren, H. Characteristics, causes and evolutionary consequences of male-biased mutation. *Proc. R. Soc. Lond. B* **274**, 1–10 (2007).
22. Goetting-Minesky, M. P. & Makova, K. D. Mammalian male mutation bias: impacts of generation time and regional variation in substitution rates. *J. Mol. Evol.* **63**, 537–544 (2006).
23. Presgraves, D. C. & Yi, S. V. Doubts about complex speciation between humans and chimpanzees. *Trends Ecol. Evol.* **24**, 533–540 (2009).
24. Patterson, N., Richter, D. J., Gnerre, S., Lander, E. S. & Reich, D. Genetic evidence for complex speciation of humans and chimpanzees. *Nature* **441**, 1103–1108 (2006).
25. Hughes, J. F. *et al.* Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature* **463**, 536–539 (2010).
26. Kamada, F. *et al.* A genome-wide association study identifies RNF213 as the first Moyamoya disease gene. *J. Hum. Genet.* **56**, 34–40 (2011).
27. Herculano-Houzel, S. Scaling of brain metabolism with a fixed energy budget per neuron: implications for neuronal activity, plasticity and evolution. *PLoS ONE* **6**, e17514 (2011).
28. Clark, A. G. *et al.* Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* **302**, 1960–1963 (2003).
29. Ellis, R. A. & Montagna, W. The skin of primates. VI. The skin of the gorilla (*Gorilla gorilla*). *Am. J. Phys. Anthropol.* **20**, 79–93 (1962).
30. Streeter, G. L. Some uniform characteristics of the primate auricle. *Anat. Rec. A* **23**, 335–341 (1922).
31. Wallis, O. C., Zhang, Y. P. & Wallis, M. Molecular evolution of GH in primates: characterisation of the GH genes from slow loris and marmoset defines an episode of rapid evolutionary change. *J. Mol. Endocrinol.* **26**, 249–258 (2001).
32. Stenson, P. D. *et al.* The Human Gene Mutation Database: 2008 update. *Genome Med.* **1**, 13 (2009).
33. Gibbs, R. A. *et al.* Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222–234 (2007).
34. Montgomery, S. B. *et al.* Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773–777 (2010).
35. Blekman, R., Marioni, J. C., Zumbo, P., Stephens, M. & Gilad, Y. Sex-specific and lineage-specific alternative splicing in primates. *Genome Res.* **20**, 180–189 (2010).
36. Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194–1211 (2009).
37. McDaniell, R. *et al.* Heritable individual-specific and allele-specific chromatin signatures in humans. *Science* **328**, 235–239 (2010).
38. Kunarso, G. *et al.* Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature Genet.* **42**, 631–634 (2010).
39. Schmidt, D. *et al.* Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328**, 1036–1040 (2010).
40. Groves, C. *Primate Taxonomy* (Smithsonian Institution Press, 2001).
41. Thalmann, O., Fischer, A., Lankester, F., Paabo, S. & Vigilant, L. The complex evolutionary history of gorillas: insights from genomic data. *Mol. Biol. Evol.* **24**, 146–158 (2007).
42. Stokes, E., Malonga, R., Rainey, H. & Strindberg, S. *Western Lowland Gorilla Surveys in Northern Republic of Congo 2006–2007. Summary Scientific Report* (WCS Global Conservation, 2008).
43. IUCN. *The IUCN Red List of Threatened Species. Version 2010. 1* (<http://www.iucnredlist.org>) (2010).
44. Stacey, M., Lin, H. H., Hilyard, K. L., Gordon, S. & McKnight, A. J. Human epidermal growth factor (EGF) module-containing mucin-like hormone receptor 3 is a new member of the EGF-TM7 family that recognizes a ligand on human macrophages and activated neutrophils. *J. Biol. Chem.* **276**, 18863–18870 (2001).
45. Jensen-Seaman, M. I. & Li, W. H. Evolution of the hominoid semenogelin genes, the major proteins of ejaculated semen. *J. Mol. Evol.* **57**, 261–270 (2003).
46. Alkan, C. *et al.* Personalized copy number and segmental duplication maps using next-generation sequencing. *Nature Genet.* **41**, 1061–1067 (2009).
47. Gazave, E. *et al.* Copy number variation analysis in the great apes reveals species-specific patterns of structural variation. *Genome Res.* **21**, 1626–1639 (2011).
48. Begun, D. R. in *Handbook of Palaeoanthropology* Vol. 2, *Primate Evolution and Human Origins* (eds Henke, W. & Tattersall, I.) 921–977 (Springer, 2007).
49. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
50. Lebatard, A. E. *et al.* Cosmogenic nuclide dating of *Sahelanthropus tchadensis* and *Australopithecus bahrelghazali*: Mio-Pliocene hominids from Chad. *Proc. Natl Acad. Sci. USA* **105**, 3226–3231 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Li and E. Birney for discussions, D. Zerbino, J. Stalker, L. Wilming, D. Rajan and H. Clawson for technical assistance, J. Ahinger for comments on the manuscript, K. Leus of the Center for Research and Conservation of the Royal Zoological Society of Antwerp for sample material from Mukisi, and the Marmoset Genome Analysis Consortium for permission to use the unpublished assembly of the marmoset genome. This research was supported in part by Wellcome Trust grants WT062023 (to J.H., K.B., S.F., A.J.V., P.F.), WT089066 (to R.D.), WT077192 (to R.D., S.M., A.K.-K., J.T.S., W.W.), WT077009 (to Y.X., B.Y., Q.A., Y.C., C.T.-S.), WT077198 (to G.K.L.) and 075491/Z/04 (to G.L.); EMBL grants (to P.C.S., P.F.); scholarships from the Gates Cambridge Trust (to G.E.J. and T.D.O'C.); an MRC Special Fellowship in Biomedical Informatics (to A.S.); funding from the Lundbeck Foundation (to A.H.); the Academy of Finland and the Emil Aaltonen Foundation (to T.L.); a Marie Curie fellowship (to T.M.-B.); the European Community's Seventh Framework Programme (FP7/2007–2013)/ERC Starting Grant (StG_20091118) (to T.M.-B.); an FPI grant from the Spanish Ministry of Education (BES-2010-032251) (to J.P.-M.); a BBSRC Doctoral Training Grant (to S.H.M.); grants from the UK Medical Research Council (to A.H., S.M., C.P.P.) and the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health (to J.C.M.); the Danish Council for Independent Research, Natural Sciences, grant no. 09-062535 (to K.M., M.H.S.); a Commonwealth Scholarship (to M.C.W.); the Swiss National Science Foundation, Louis Jeantet Foundation (to E.T.D.); an ERC Starting Grant and an EMBO Young Investigator Award, Hutchinson Whampoa (to D.T.O.); NHGRI support (to W.C.W.); support from BIOBASE GmbH (to E.V.B., P.D.S., M.M., A.D.P., K.S., D.N.C.); US National Science Foundation grant DGE-0739133 (to W.W.); NHGRI U54 HG003079 (to R.K.W.); NIH grant HG002385 (to E.E.E.). E.E.E. is an investigator of the Howard Hughes Medical Institute.

Author Contributions Manuscript main text: A.S., R.D., C.T.-S., N.I.M., G.E.J., P.C.S., A.K.-K. Project coordination: A.S., A.S.R., A.K.-K., R.D. Project initiation: J.R., R.D., R.K.W. Library preparation and sequencing: I.G., D.J.T., M.A.Q., C.M.C., B.Z., P.d.J., O.A.R., Q.A., B.Y., Y.X., T.A.G., W.C.W. Assembly: A.S., L.W.H., Y.G., J.T.S., J.C.M., W.W., Z.N. Fosmid finishing: P.H. Assembly quality: A.S., S. Meader, G.L., C.P.P. Annotation: Y.A.T., G.K.L., A.J.V., A. Heger, S.M.S. Primate multiple alignments: J.H., K.B., S.F. Great ape speciation and ILS: J.Y.D., A.S., T.M., M.H.S., K.M., G.E.J. Sequence loss and gain: A.S., S.M., C.T.-S., Y.A.T., A.J.V. Protein evolution: G.E.J., S.H.M., N.I.M., B.J.B., T.D.O'C., Y.X., Y.C., N.G. Human disease allele analysis: Y.X., Y.C., C.T.-S., P.D.S., E.V.B., A.D.P., M.M., K.S., D.N.C. Transcriptome analysis: T.L., E.T.D. ChIP-seq experiment and analysis: P.C.S., M.C.W., D.S., P.F., D.T.O. Additional gorilla samples: B.Y., Y.X., L.V., C.T.-S. Gorilla species diversity and divergence: A.S., A.H., T.M., L.N.A., B.Y., L.V. Gorilla species functional differences: Y.X., Y.C., C.T.-S. Segmental duplication analysis: T.M.-B., C.A., S.S., E.K., J.P.-M., E.E.E.

Author Information Accession numbers for all primary sequencing data are given in Supplementary Information. The assembly has been submitted to EMBL with accession numbers FR853080 to FR853106, and annotation is available at Ensembl (http://www.ensembl.org/Gorilla_gorilla/Info/Index). Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.D. (rd@sanger.ac.uk).

Type VI secretion requires a dynamic contractile phage tail-like structure

M. Basler^{1*}, M. Pilhofer^{2,3*}, G. P. Henderson², G. J. Jensen^{2,3} & J. J. Mekalanos¹

Type VI secretion systems are bacterial virulence-associated nanomachines composed of proteins that are evolutionarily related to components of bacteriophage tails. Here we show that protein secretion by the type VI secretion system of *Vibrio cholerae* requires the action of a dynamic intracellular tubular structure that is structurally and functionally homologous to contractile phage tail sheath. Time-lapse fluorescence light microscopy reveals that sheaths of the type VI secretion system cycle between assembly, quick contraction, disassembly and re-assembly. Whole-cell electron cryotomography further shows that the sheaths appear as long tubular structures in either extended or contracted conformations that are connected to the inner membrane by a distinct basal structure. These data support a model in which the contraction of the type VI secretion system sheath provides the energy needed to translocate proteins out of effector cells and into adjacent target cells.

Secretion systems allow bacteria to transport macromolecules such as proteins out of effector cells or into either target host cells during pathogenesis or target bacterial cells during competition in various ecological settings. The type 6 secretion systems (T6SS) are encoded by a cluster of 15–20 genes that is present in at least one copy in about 25% of all sequenced Gram-negative bacteria. Although linked to virulence during host infection, recent studies showed that T6SS of *Pseudomonas*, *Burkholderia* and *Vibrio* species can kill bacterial as well as eukaryotic target host cells^{1–7}. T6SS⁺ bacterial effector cells are thought to kill target cells through translocation of toxic effector proteins in a cell–cell contact-dependent process^{1–3,8}. Little is known, however, about how T6SS transport toxic proteins through their own cell membranes or across target cell membranes.

Bioinformatic⁹ and structural analyses^{10,11} have indicated that some T6SS components are structural homologues of components present in contractile phage tails. For example, secreted VgrG proteins are structural homologues of the T4 phage needle or spike complex^{9,10} and secreted Hcp is a structural homologue of a phage tail tube protein^{10,11}. Another highly conserved T6SS gene product is predicted to be a homologue of gp25, a major component of the T4 phage tail baseplate^{10,12,13}. Two T6SS gene products of *V. cholerae*, VipA and VipB, form tubular structures that can be depolymerized by another T6SS gene product, ClpV^{14–16}. Leiman *et al.*¹⁰ noted that VipA/VipB tubules visually resemble T4 contracted tail sheath and were the first to propose that a sheath-like structure might power T6SS translocation by a phage tail-like contraction mechanism. Here we show that T6SS-dependent secretion of Hcp and killing of *Escherichia coli* by *V. cholerae* correlates with the activity of a dynamic intracellular structure that indeed appears structurally and functionally related to contractile phage tail sheath.

Fluorescence microscopic imaging of the T6SS

To test the hypothesis that the T6SS apparatus is a dynamic structure, we constructed a carboxy (C)-terminal fusion of VipA protein with superfolder green fluorescent protein (sfGFP)¹⁷. As shown in Supplementary Fig. 1a, VipA–sfGFP complements a chromosomal in-frame deletion

of *vipA* for Hcp secretion when the fusion protein is expressed from pBAD24 plasmid to the same level as the wild-type allele. Visualization of cells expressing the functional sfGFP fusion protein by fluorescence microscopy revealed that the VipA–sfGFP fusion is associated with long straight structures in the cytosol that spread throughout the width and length of the cell. The number of visible structures in a single cell varied from 0 to 5 in wild-type background cells (Supplementary Fig. 2a). Critically, these structures were not visible in *vipB* mutant cells (Supplementary Fig. 2b), suggesting that the fluorescent structures could be the T6SS sheath structures hypothesized to form in part by interaction of VipA with VipB¹⁰. Because VipA is not secreted and resides within cellular fractions (Supplementary Fig. 3) and the expression level of VipA–sfGFP was comparable to VipA under the conditions used to visualize these sheath structures (Supplementary Fig. 4), we conclude that the fluorescent structures were within the cytosol of imaged cells.

Time-lapse imaging revealed these putative sheath structures to be highly dynamic. As shown in Fig. 1a and Supplementary Videos 1 and 2, the VipA–sfGFP labelled sheaths extended within tens of seconds in different subcellular locations and then contracted and disassembled also within tens of seconds. Most of the extended sheath structures visible in cells stretched from one lateral side of the cell to the other, perpendicular to the membrane, and thus had lengths approximately equal to the width of the cell (about 0.75–1 μm). As shown in Fig. 1a, these sheaths assembled at speeds of approximately 20–30 s^{-1} . Contraction was very fast, occurring in approximately 5 ms or less (unresolvable at frame rates of approximately 200 frames per second; see Fig. 1b, c, Supplementary Fig. 5 and Supplementary Video 3). Sheaths contracted to about 50% of their extended length (Supplementary Fig. 6), and then disassembled over the next 30–60 s (Fig. 1a). The disassembly of the contracted sheath is most probably a ClpV-dependent event because ClpV is known to disassemble VipA/VipB tubules *in vitro* in the presence of ATP¹⁴ and *clpV* mutants do not disassemble VipA–sfGFP-labelled contracted sheaths (Supplementary Fig. 2d and Supplementary Video 8). A similar number of VipA–sfGFP-labelled sheaths were seen in strains 2740-80 and V52 and at various levels of VipA–sfGFP expression (Supplementary

¹Department of Microbiology and Immunobiology, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA. ²Division of Biology, California Institute of Technology, 1200 East California Boulevard, Pasadena, California 91125, USA. ³Howard Hughes Medical Institute, California Institute of Technology, 1200 East California Boulevard, Pasadena, California 91125, USA.

*These authors contributed equally to this work.

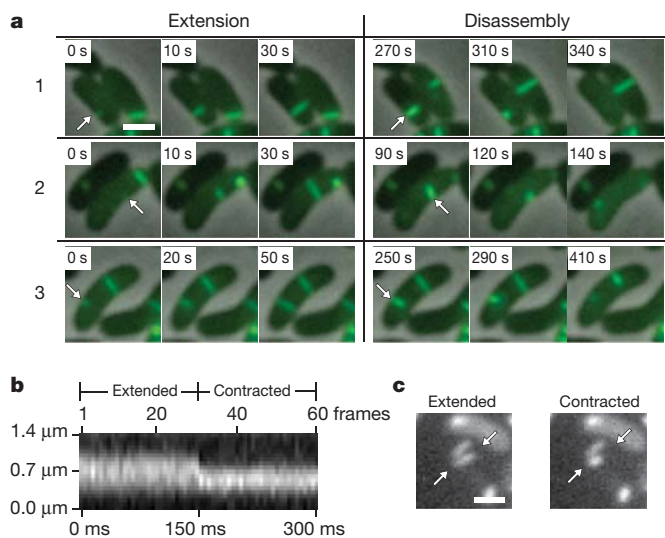


Figure 1 | Fluorescence light microscopy of VipA-sfGFP. **a**, Individual $3\ \mu\text{m} \times 3\ \mu\text{m}$ frames from a time-lapse imaging with a frame rate of 10 s per frame show three frames of extension of VipA-sfGFP structure in ΔVipA background from one side of the cell to another (arrows) followed by a contraction event and apparent disassembly (shown on three frames) of the contracted VipA-sfGFP structure (arrows). Scale bar on the first frame represents $1\ \mu\text{m}$. The whole 10 min time-lapse sequence is shown in Supplementary Video 1 with another 17 similar events; a larger field of cells is shown in Supplementary Video 2. **b**, Kymogram illustrating rapid change in the length of the VipA-sfGFP structure. Projection of signal intensity in time at a rate of 200 frames per second along the axis of the maximal intensity on an extended structure (30 frame average shown on **c** left) showing a contraction in length and increase in maximal intensity of the contracted structure (30 frame average shown on **c** right). Arrows indicate contracting VipA-sfGFP structure and mark the start and end of a line for generating the kymogram. Scale bar shown on **c** left represents $1\ \mu\text{m}$. Gaussian blur filter (sigma radius = 1) was applied to individual frames before generating the kymogram. All 60 frames of the time-lapse sequence are shown in Supplementary Video 3 (video segment number 3) with four more contraction events imaged at the same or lower frame rate.

Figs 2c, g and 4 and Supplementary Videos 4–6). The sheaths in wild-type cells displayed the same extension–contraction–disassembly cycles as sheaths observed in complemented *vipA* mutant cells (Supplementary Videos 4 and 5) and when mCherry2 (ref. 18) was substituted for sfGFP (Supplementary Fig. 2f and Supplementary Video 7). Thus, the dynamic behaviour observed is not just a property of sheaths that contain exclusively VipA-sfGFP fusion protein, but also of sheaths composed largely of wild-type VipA or other VipA fusion proteins. VCA0109 encodes a member of a family of phage base-plate proteins¹⁰. In a VCA0109 deletion background most of the VipA-sfGFP

was dispersed in the cytosol with only a rare, VipA-sfGFP-containing structure ever visible (Supplementary Fig. 2e and Supplementary Video 9). VCA0109 therefore plays a critical role in the formation of functional T6SS sheaths.

Electron cryotomographic imaging of the T6SS

To visualize the sheaths directly, we imaged wild-type and mutant whole cells with electron cryotomography (ECT). ECT has been shown to preserve and reveal bacterial cytoskeletal structures directly in three dimensions in a near-native, life-like state¹⁹. ECT analyses of wild-type 2740-80 cells showed straight, tubular structures which

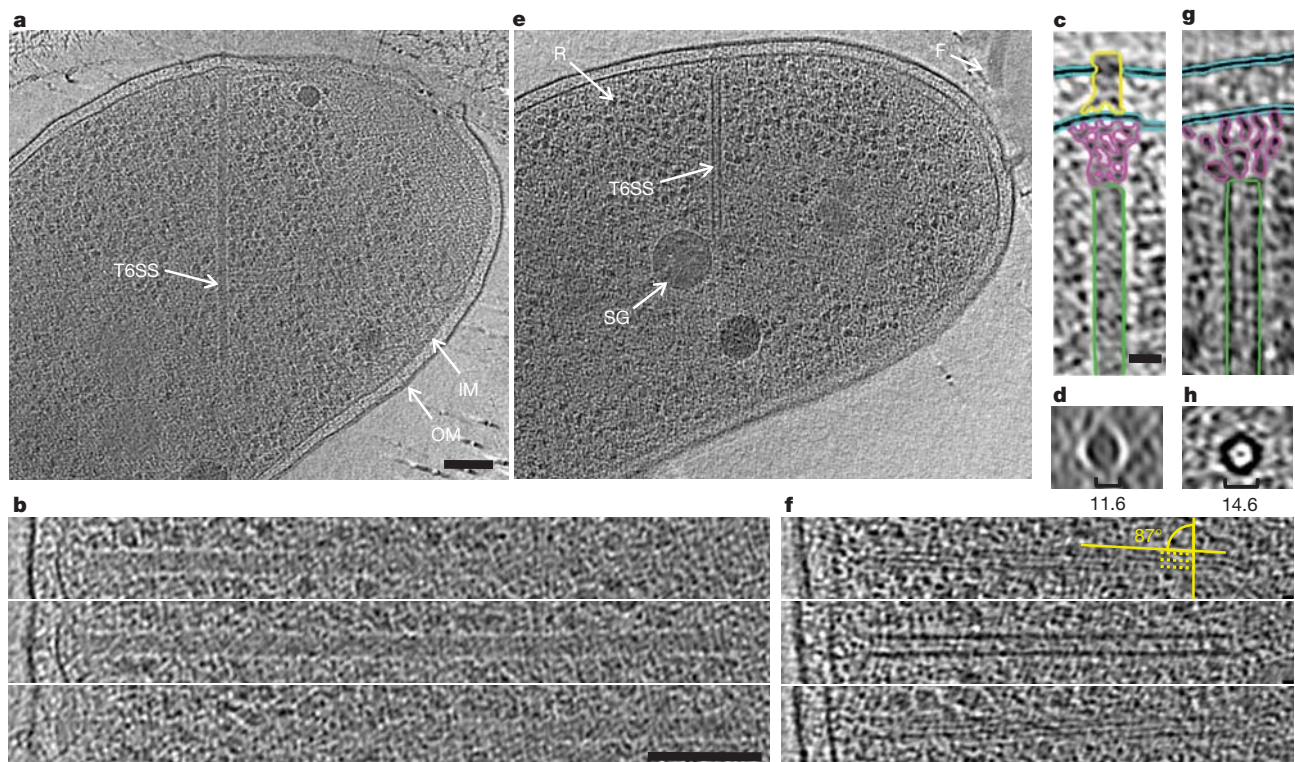


Figure 2 | Electron cryotomographic imaging of T6SS structures inside intact cells. Shown are different tomographic slices (19 nm in **a**, **e**, **c**, **g**; 9.5 nm in **b**, **f**; 190 nm in **d**, **h**) of an extended (**a**–**d**) and a contracted (**e**–**h**) structure imaged in two different wild-type cells (contracted/extended structures, T6SS; IM, inner membrane; OM, outer membrane; F, flagellum; R, putative ribosome; SG, polyphosphate storage granule). **b**, **f**, Each part shows three slices at the same orientation but at different z-heights. Compared with extended structures, contracted structures are shorter (**b**, **f**), have a helical surface pattern

(pitch angle 87°) and a smaller diameter (indicated in the perpendicular views in **d**, **h**). **c**, **g**, Segmentations of densities observed in the extended (**c**) and contracted (**g**) structures. Densities shown in **h** originate from a contracted structure from a different tomogram. Segmented are putative densities corresponding to sheath (green), base plate (pink and yellow) and membranes (blue). Scale bars: **a**, 100 nm (applies to **a**, **e**); **b**, 100 nm (applies to **b**, **f**); **c**, 20 nm (applies to **c**, **d**, **g**, **h**).

appeared to exist in two different conformations: a longer, thinner 'extended' conformation (Fig. 2a–d) and a shorter, wider, 'contracted' conformation (Fig. 2e–h). Both structures were oriented roughly perpendicular to the cytoplasmic membrane and were clearly located exclusively in the cytosol (Supplementary Video 10). Tubular structures were observed in 26 of 90 imaged wild-type cells. Some cells exhibited more than one tubule and on occasion both extended and contracted conformations were seen in the same cell (Supplementary Table 1 and Supplementary Fig. 7). No tubular structures were observed in a *vipB* mutant (0 of 53 cells), a VCA109 mutant (0 of 10 cells) and a VCA0109/ClpV double mutant (0 of 8 cells), strongly suggesting that both types of tubule are T6SS-related structures.

Consistent with the dynamic sheaths in the two-dimensional fluorescence projection images, the extended and contracted tubes seen in the three-dimensional cryotomograms had lengths of 667 ± 83 nm ($n = 13$) and 372 ± 56 nm ($n = 16$), respectively. Although extended tubes had diameters of 11.6 ± 0.7 nm, dense interiors, and a homogeneous surface, we observed that contracted tubes were thicker (14.6 ± 0.7 nm diameter), hollow, and had helical ridges (87° pitch angle) spaced 6 nm apart (Fig. 2). The tubular structures of both conformations were usually found with one end in close proximity to the cytoplasmic membrane in a near-perpendicular orientation (Fig. 2 and Supplementary Fig. 8). The tubes did not contact the membrane directly, however, but instead appeared to be connected to it by a flared bell-shaped base (Fig. 2c, g, pink highlights). Distal to the flared base of extended, but not contracted tubes, there was an additional conical-shaped density (Fig. 2c, yellow highlights) that crossed the periplasm and protruded through the outer membrane. Given that various T6SS components have been localized to the inner membrane, periplasm and outer membrane including a lipoprotein unique to T6SS as well as proteins related to orthologues IcmF, DotU, and OmpA in other organisms^{20–22}, we propose that all these densities simply be called the 'T6SS base plate' complex.

Purification of T6SS sheath from *V. cholerae*

To prove that the dynamic fluorescent structures observed in VipA–sfGFP expressing cells and the tubes observed by ECT were indeed T6SS sheath-related structures, we sought to purify the corresponding structures from disrupted cells. Negative stain electron microscopic analysis of macromolecular fractions purified from cell lysates of wild-type cells revealed straight, hollow tubular structures similar to, but more uniform than, the VipA/VipB tubules produced previously in *E. coli*¹⁴ and distinct from the *V. cholerae* flagellum (Fig. 3a left and Supplementary Fig. 1b left). No sheath-like structures were detected in identically prepared samples from mutants defective in VipA or VipB, although flagella were clearly present (data not shown). Electron microscopic analysis of sheath preparations made from the VipA–sfGFP complemented *vipA* mutant strain revealed sheath structures similar to those produced by wild-type cells except that a diffuse coat was laterally displayed on the filament's surface, probably corresponding to the sfGFP moiety on the fusion protein (Fig. 3a right and Supplementary Fig. 1b right).

To identify proteins that were associated with these sheaths, the structures were purified from a non-flagellated mutant (*flgG*) of *V. cholerae* 2740–80 (Supplementary Fig. 9a). Two major proteins were enriched in these preparations with apparent molecular masses of 55 and 20 kDa, respectively (Supplementary Fig. 9b). Mass spectrometry analysis revealed that the 55 kDa band was VipB and the 20 kDa band was VipA (Supplementary Table 2). Interestingly, we also identified four additional T6SS proteins in the sheath samples: ClpV, VCA0109 (a gp25-like protein), and two other proteins within the T6SS gene cluster encoded by genes *VCA0111* and *VCA0114*. ClpV was recently shown to interact directly with VipB, most strongly in its polymerized state with VipA¹⁵. As noted earlier, VCA0109 is a homologue of T4 base-plate protein gp25 (ref. 10) and a T6SS gp25-like protein was recently shown to localize to the cytoplasm of

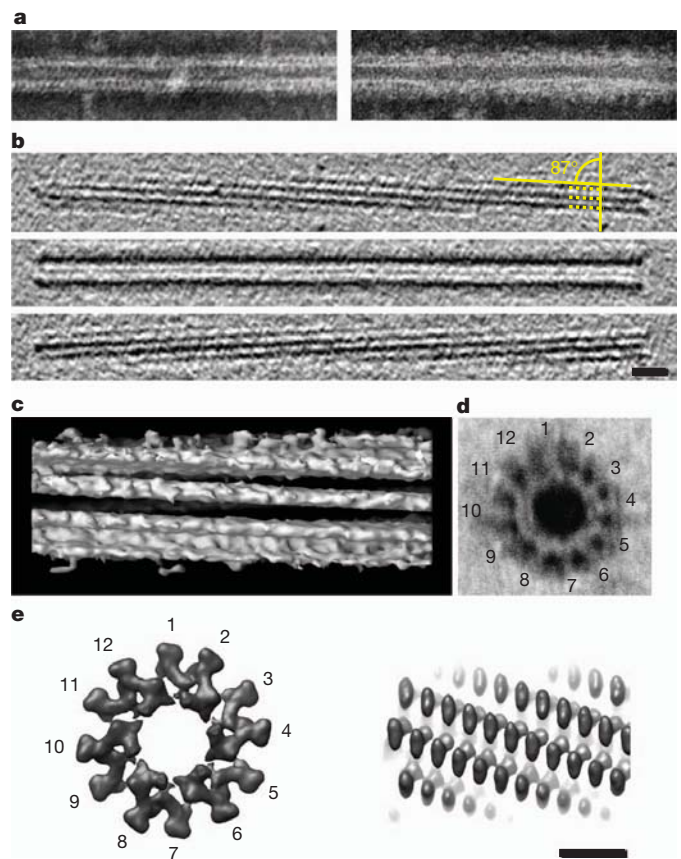


Figure 3 | Images of purified VipA/VipB sheaths and comparison with phage tails. Negative staining (a) images of purified wild-type sheath (left) and VipA–sfGFP-labelled sheath (right) are highly similar except for flared extra densities on the outside of the VipA–sfGFP-labelled structure. Cryotomograms of wild-type sheath (b, showing three 12.6-nm slices at different z -heights) were highly similar to contracted structures imaged *in vivo* (Fig. 2f). Note the matching surface pitch angle of 87° seen in tomographic slices (b) and an isosurface of a subtomogram average (c; map deposited in the Electron Microscopy Data Base under accession EMD-2045). The negatively stained perpendicular view of a purified wild-type sheath showed the characteristic 'cog-wheel' like structure with 12 paddles (d) and is similar to the perpendicular view of a contracted T4 phage sheath (e, left; two rings of six gp18 subunits, created in Chimera from EMDB 1086 map). Similar to T6SS sheath (c), also the surface of the contracted T4 phage sheath appears helical (e, right) though with a different pitch angle. Scale bars: b, 20 nm (applies to a, b); e, 10 nm (applies to c–e). Note that protein densities appear white in negative stain images and black in cryotomograms.

*Pseudomonas aeruginosa*¹³. The function of VCA0111 and VCA0114 are currently unknown, but they are essential components of the *V. cholerae* T6SS⁷.

The VipA/VipB sheath preparations purified from wild-type 2740–80 cells were also imaged by ECT (Fig. 3b, c), by which they were recognized as the contracted tubes seen previously inside cells (14.4 nm diameter, hollow interior, and helical surface ridges spaced 6 nm apart with pitch angle 87°). Interestingly, in addition to helical surface ridges, purified sheaths exhibited cogwheel-like cross-sections with 12 surface 'paddles' per rotation (Fig. 3d, seen most clearly after negative staining) and thus are structurally similar to contracted T4 phage sheaths¹² (Fig. 3e and Supplementary Fig. 10). We conclude that the two tubular structures seen *in vivo* by ECT correspond to extended and contracted states of the dynamic VipA–sfGFP-labelled T6SS sheath that was visualized using fluorescence light microscopy.

Contractile phage tails consist of a contractile outer sheath and an inner tube that is projected into a target host cell during phage infection²³. The T6SS of *V. cholerae* is known to possess antibacterial activity against *E. coli* that correlates with its ability to secrete the T6SS substrate

protein Hcp^{4,7}. As shown in Supplementary Figs 1a and 11a, *V. cholerae* strain 2740-80 secretes abundant Hcp and this secretion is completely abolished by deletion of T6SS genes *vipA*, *vipB* and *VCA0109* (which encodes a gp25-like protein), as has been previously shown for *V. cholerae* strain V52 (refs 7, 14). Although the material inside the extended tubule visualized with ECT (Fig. 2d and Supplementary Fig. 10a) could not be resolved as a separate ‘inner tube’ per se, its diameter was similar to the diameter of Hcp tubes described at either the crystallographic or microscopic level^{6,24}. Furthermore, contracted tubes were clearly hollow (Fig. 2h and Supplementary Fig. 10c–f). Thus, we propose that the thinner extended tubule found in whole cells is an uncontracted ‘extended T6SS sheath’ whose VipA/VipB subunits are probably wrapped around a thinner inner tube composed of Hcp protein. Unfortunately, the uncontracted, extended T6SS sheaths could not be purified from *V. cholerae* cells for further analysis, perhaps because of spontaneous sheath contraction during cellular disruption and purification. Because Hcp was not found in purified contracted T6SS sheaths, we conclude that the postulated inner Hcp tube of extended sheaths is largely expelled from the cell at the moment of contraction.

ClpV and T6SS sheath recycling

Like strain V52 (refs 4, 7), *V. cholerae* 2740-80 also rapidly kills *E. coli* when co-cultivated on agar in a *vipA*-, *vipB*-, and *VCA0109*-dependent

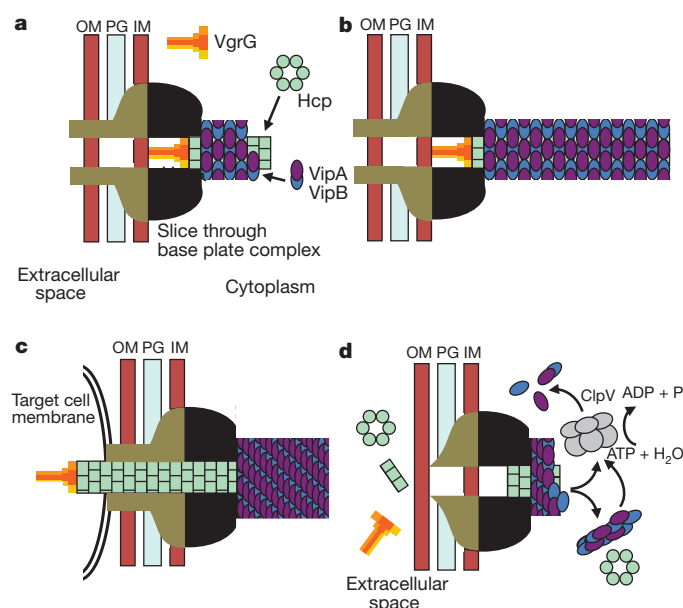


Figure 4 | Model of T6SS action. OM, outer membrane; PG, peptidoglycan; IM, inner membrane. **a**, Assembly. The first step is a base-plate complex formation that initiates the Hcp tube polymerization. The base-plate complex is probably composed of gp25, VgrG and other T6SS proteins that define a bell-shaped cytoplasmic component (black objects) and periplasmic component (brown objects), which together span the inner membrane, peptidoglycan and outer membrane. The second step is polymerization of the sheath (from VipA/VipB heterodimers) around the Hcp tube in an extended conformation. **b**, Extended T6SS apparatus in extended ‘ready to fire’ conformation. The membrane distal end may be capped by an unknown protein or VipAB conformational state. **c**, Contraction. Upon an unknown extracellular signal, a conformational change in the base-plate complex triggers sheath contraction that leads to the translocation (secretion) of the VgrG/Hcp tube complex through effector cell membranes and penetration of adjacent target cell membrane. Translocation of additional effector proteins might then follow using the Hcp tube as a conduit. **d**, Disassembly. Contracted sheath is detached and disassembled by ClpV ATPase. VipA/B dimers released are recycled into a new extended T6SS apparatus at either the original or a newly formed base-plate complex. In the absence of target cell penetration (c), Hcp and VgrG proteins are released into the extracellular space as secreted proteins.

fashion (Supplementary Fig. 11b). Consistent with published results in strain V52 (ref. 7), the ClpV mutant of 2740-80 showed 90% loss of T6SS-dependent killing in 2 h assays but retained bacteriocidal activity well above background during incubation for 4 h. Thus, ClpV is not essential for T6SS function in *V. cholerae*. Because ClpV has been shown to disassemble *in vitro* a tubular structure that is produced in *E. coli* expressing VipA and VipB¹⁴, we asked whether ClpV affected the dynamics of T6SS sheath imaged with the VipA–sfGFP fusion. We did not observe any polymerization or disassembly events in the ClpV mutant background; rather, we found that most VipA–sfGFP existed in static punctate structures (Supplementary Fig. 2d), which were probably contracted T6SS sheaths because ClpV mutants produce contracted sheath-like structures (Supplementary Fig. 1b).

Because our fluorescence microscopic analysis showed that contracted sheath forms from extended sheath, it follows that ClpV may not play a role in T6SS sheath assembly or secretion function; rather, it is responsible for recycling VipA and VipB from contracted T6SS sheath structures through disassembly. The ClpV-mediated disassembly process frequently begins with dislocation of the contracted sheath from the original site of extension (that is, the T6SS base-plate complex) and then continued disassembly in a random fashion throughout the cell cytosol; VipA–sfGFP released by this process is quickly reassembled into new extended sheaths in many cells (Supplementary Videos 1 and 2). Based on these observations we propose a detailed model of the dynamic steps in the functional cycle of the T6SS apparatus that will serve as a framework for further studies (Fig. 4 and Supplementary Video 11, a narrated animation of our data and model). Although analogous to translocation events mediated by contractile phage tails, the proposed T6SS process is different because it occurs in a topologically reversed orientation and compartmentalization (within the cytosol), and further undergoes efficient recycling through the action of other T6SS components such as ClpV. Collectively, the data presented here provide strong evidence that energy captured from conformational changes in polymeric structures can rapidly transport proteins through cell membranes.

METHODS SUMMARY

V. cholerae strains and genetic manipulation have been described previously^{7,25,26}. Full-length VipA was fused at its C terminus to either sfGFP or mCherry2 genes (separated by a 3×Ala 3×Gly linker) and expressed from an arabinose-inducible promoter²⁷ present on either plasmid pBAD24–VipA–sfGFP or pBAD24–VipA–mCherry2. Cells grown to an attenuation at 600 nm (D_{600}) of 1.0 in the presence of various concentrations of arabinose were spotted on a thin pad of 1% agarose in PBS and imaged at room temperature. Fluorescence micrographs were captured using microscopes and camera combinations that are described in detail in Methods and image analysis was performed using ImageJ 1.45 software. For electron microscopy, sheath samples prepared as described in Methods were spotted on carbon-coated grids, and stained with 1% uranyl formate. The grids were examined in a JEOL 1200EX transmission electron microscope and images were recorded with an AMT 2k CCD (charge-coupled device) camera. For ECT, *V. cholerae* cells were grown to D_{600} = 1.5–2.2, mixed with 10 nm colloidal gold, applied to an electron microscope grid and plunge-frozen in a liquid ethane–propane mixture²⁸. Tilt series were collected using a Polara 300 kV FEG transmission electron microscope equipped with an energy filter (slit width 20 eV) on a lens-coupled 4k × 4k UltraCam. Pixels on the CCD represented 0.95 nm (×22,500) or 0.63 nm (×34,000) at the specimen level. Legimon²⁹ or UCSF Tomo³⁰ was used for automatic tilt-series acquisition. Three-dimensional reconstructions were calculated using the IMOD software package³¹ or Raptor³². IMOD³¹ was used to model the centre of the sheath, PEET³³ to align and average repeating sub-volumes, and Chimera³⁴ to do isosurface rendering of the sub-volume averages.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 October 2011; accepted 9 January 2012.

Published online 26 February 2012.

1. Pukatzki, S. *et al.* Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system. *Proc. Natl Acad. Sci. USA* **103**, 1528–1533 (2006).

2. Ma, A. T., McAuley, S., Pukatzki, S. & Mekalanos, J. J. Translocation of a *Vibrio cholerae* type VI secretion effector requires bacterial endocytosis by host cells. *Cell Host Microbe* **5**, 234–243 (2009).
3. Russell, A. B. *et al.* Type VI secretion delivers bacteriolytic effectors to target cells. *Nature* **475**, 343–347 (2011).
4. MacIntyre, D. L., Miyata, S. T., Kitaoka, M. & Pukatzki, S. The *Vibrio cholerae* type VI secretion system displays antimicrobial properties. *Proc. Natl Acad. Sci. USA* **107**, 19520–19524 (2010).
5. Schwarz, S. *et al.* *Burkholderia* type VI secretion systems have distinct roles in eukaryotic and bacterial cell interactions. *PLoS Pathog.* **6**, e1001068 (2010).
6. Hood, R. D. *et al.* A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria. *Cell Host Microbe* **7**, 25–37 (2010).
7. Zheng, J., Ho, B. & Mekalanos, J. J. Genetic analysis of anti-amoebae and anti-bacterial activities of the type VI secretion system in *Vibrio cholerae*. *PLoS ONE* **6**, e23876 (2011).
8. Ma, A. T. & Mekalanos, J. J. *In vivo* actin cross-linking induced by *Vibrio cholerae* type VI secretion system is associated with intestinal inflammation. *Proc. Natl Acad. Sci. USA* **107**, 4365–4370 (2010).
9. Pukatzki, S., Ma, A. T., Revel, A. T., Sturtevant, D. & Mekalanos, J. J. Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proc. Natl Acad. Sci. USA* **104**, 15508–15513 (2007).
10. Leiman, P. G. *et al.* Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc. Natl Acad. Sci. USA* **106**, 4154–4159 (2009).
11. Pell, L. G., Kanelis, V., Donaldson, L. W., Howell, P. L. & Davidson, A. R. The phage lambda major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. *Proc. Natl Acad. Sci. USA* **106**, 4160–4165 (2009).
12. Leiman, P. G., Chipman, P. R., Kostyuchenko, V. A., Mesyanzhinov, V. V. & Rossmann, M. G. Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. *Cell* **118**, 419–429 (2004).
13. Lossi, N. S., Dajani, R., Freemont, P. & Filloux, A. Structure-function analysis of HsfI, a gp25-like component of the type VI secretion system in *Pseudomonas aeruginosa*. *Microbiology* **157**, 3292–3305 (2011).
14. Bonemann, G., Pietrosiuk, A., Diemand, A., Zentgraf, H. & Mogk, A. Remodelling of VipA/VipB tubules by ClpV-mediated threading is crucial for type VI protein secretion. *EMBO J.* **28**, 315–325 (2009).
15. Pietrosiuk, A. *et al.* Molecular basis for the unique role of the AAA⁺ chaperone ClpV in type VI protein secretion. *J. Biol. Chem.* **286**, 30010–30021 (2011).
16. Mougous, J. D. *et al.* A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science* **312**, 1526–1530 (2006).
17. Pedelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnol.* **24**, 79–88 (2006).
18. Cho, H., McManus, H. R., Dove, S. L. & Bernhardt, T. G. Nucleoid occlusion factor SlmA is a DNA-activated FtsZ polymerization antagonist. *Proc. Natl Acad. Sci. USA* **108**, 3773–3778 (2011).
19. Pilhofer, M., Ladinsky, M. S., McDowell, A. W. & Jensen, G. J. Bacterial TEM: new insights from cryo-microscopy. *Methods Cell Biol.* **96**, 21–45 (2010).
20. Aschtgen, M. S., Bernard, C. S., De Bentzmann, S., Llobes, R. & Cascales, E. SciN is an outer membrane lipoprotein required for type VI secretion in enteroaggregative *Escherichia coli*. *J. Bacteriol.* **190**, 7523–7531 (2008).
21. Aschtgen, M. S., Thomas, M. S. & Cascales, E. Anchoring the type VI secretion system to the peptidoglycan: TssL, TagL, TagP, what else? *Virulence* **1**, 535–540 (2010).
22. Aschtgen, M. S., Gavioli, M., Dessen, A., Llobes, R. & Cascales, E. The SciZ protein anchors the enteroaggregative *Escherichia coli* type VI secretion system to the cell wall. *Mol. Microbiol.* **75**, 886–899 (2010).
23. Kostyuchenko, V. A. *et al.* The tail structure of bacteriophage T4 and its mechanism of contraction. *Nature Struct. Mol. Biol.* **12**, 810–813 (2005).
24. Ballister, E. R., Lai, A. H., Zuckermann, R. N., Cheng, Y. & Mougous, J. D. *In vitro* self-assembly of tailorable nanotubes from a simple protein building block. *Proc. Natl Acad. Sci. USA* **105**, 3733–3738 (2008).
25. Goldberg, S. & Murphy, J. R. Molecular epidemiological studies of United States Gulf Coast *Vibrio cholerae* strains: integration site of mutator vibriophage VcA-3. *Infect. Immun.* **42**, 224–230 (1983).
26. Bina, J. E. & Mekalanos, J. J. *Vibrio cholerae* tolC is required for bile resistance and colonization. *Infect. Immun.* **69**, 4681–4685 (2001).
27. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).
28. Tivol, W. F., Briegel, A. & Jensen, G. J. An improved cryogen for plunge freezing. *Microsc. Microanal.* **14**, 375–379 (2008).
29. Suloway, C. *et al.* Fully automated, sequential tilt-series acquisition with Leginon. *J. Struct. Biol.* **167**, 11–18 (2009).
30. Zheng, S. Q. *et al.* UCSF tomography: an integrated software suite for real-time electron microscopic tomographic data collection, alignment, and reconstruction. *J. Struct. Biol.* **157**, 138–147 (2007).
31. Mastronarde, D. N. Correction for non-perpendicularity of beam and tilt axis in tomographic reconstructions with the IMOD package. *J. Microsc.* **230**, 212–217 (2008).
32. Amat, F. *et al.* Markov random field based automatic image alignment for electron tomography. *J. Struct. Biol.* **161**, 260–275 (2008).
33. Nicastro, D. *et al.* The molecular architecture of axonemes revealed by cryoelectron tomography. *Science* **313**, 944–948 (2006).
34. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. G. Bernhardt and N. T. Peters for assistance with fluorescence microscopy, discussions and for a gift of plasmids carrying sfGFP and mCherry2 genes. We thank the Nikon Imaging Center at Harvard Medical School for help with fluorescence microscopy, and Research Precision Instruments and Hamamatsu for lending an ORCA-Flash2.8 camera. We thank the Harvard Medical School Electron Microscopy Facility for help with and supervision of transmission electron microscopy. We thank M. K. Waldor for a *V. cholerae* 2740-80 strain and discussions. We thank D. Ewen Cameron for a knockout construct pWM91-flgG. We thank B. Wen and Z. Li for initial cryotomographic studies. This work was supported by National Institute of Allergy and Infectious Diseases grants AI-018045 and AI-26289 to J.J.M. and National Institute of General Medical Sciences grant GM094800B to G.J.J.

Author Contributions All authors helped design and analyse experiments; M.B., M.P. and G.P.H. performed experiments, and M.B., M.P., G.J.J. and J.J.M. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.J.J. (Jensen@caltech.edu) or J.J.M. (john_mekalanos@hms.harvard.edu).

METHODS

Bacterial strains. *V. cholerae* 2740-80 is a non-toxinogenic El Tor strain isolated in 1980 from a patient in Florida, USA²⁵. A streptomycin resistant, *lacZ*⁻ derivative of 2740-80 was used as the wild-type parental strain. *E. coli* DH10 β and Sm10 λ pir were used for cloning and conjugation, respectively. Gentamicin-resistant *E. coli* MG1655 strain was used in bacterial killing assays. *V. cholerae* V52 and its deletion variants were described previously⁷. Antibiotic concentrations used were streptomycin (100 μ g ml⁻¹), gentamicin (15 μ g ml⁻¹) and carbenicillin (100 μ g ml⁻¹). Luria–Bertani (LB) broth was used for all growth conditions. Liquid cultures were grown aerobically at 37 °C.

DNA manipulations. To generate an in-frame deletion in *vipA*, *vipB*, *clpV*, *flgG* or VCA0109, the corresponding surrounding DNA was amplified by overlap extension PCR and cloned into pWM91 (ref. 35) for subsequent *sacB*-mediated allelic exchange as described²⁶. Primers were designed such that each deletion resulted in the replacement the entire open reading frame, with the exception of first and last seven codons. Gene deletion was confirmed by PCR with primers outside of the replaced region. For construction of variants of pBAD24, full-length *vipA* gene was amplified from chromosomal DNA and sfGFP or mCherry2 genes were amplified from plasmids carrying respective genes. Full-length *vipA* gene, or *vipA* genes that had been fused with either sfGFP or mCherry2 genes (both separated by a DNA linker encoding 3 \times Ala 3 \times Gly), were cloned into plasmid pBAD24 (ref. 27). All cloning products were sequence-verified.

Bacterial killing assay. *V. cholerae* 2740-80 strains and *E. coli* MG1655 strain were incubated for 14–18 h at 37 °C in LB, then washed in fresh LB and diluted $\times 10$ in LB. Attenuance at 600 nm (D_{600}) of the culture was adjusted to 0.4 for both *V. cholerae* and *E. coli*. *V. cholerae* was mixed with *E. coli* in a 10:1 ratio and 10 μ l of the mixture was spotted on a dry LB agar plate. After 2 and 4 h, bacterial spots were cut out and the cells were re-suspended in 0.5 ml LB. The cellular suspension was serially diluted in LB and 5 μ l of the suspensions was spotted on selective plates (gentamicin for *E. coli*, streptomycin 100 μ g ml⁻¹ *V. cholerae*). Colonies were detected after incubation for approximately 16 h at 30 °C.

Cell fractionation and immunoblot analysis. Cells from overnight cultures were washed with fresh LB and diluted 1:100 in 1.5 ml of fresh LB (supplemented with appropriate antibiotics and arabinose to indicated concentrations), cultivated for 2.5–3.0 h to D_{600} about 1.0. Cells were collected by centrifugation at 21,000g for 1 min and re-suspended in 250 μ l SDS–PAGE loading buffer; 15 μ l was loaded for western blot analysis. Cell-free supernatants (1.0 ml) were precipitated by 10% trichloroacetic acid for 1 h on ice. Precipitated proteins were collected by centrifugation for 15 min at 21,000g, washed with 100% acetone and re-suspended in 60 μ l SDS–PAGE loading buffer. Twenty microlitres was loaded on an SDS–PAGE for western blot analysis. Cell and supernatant protein samples were boiled for 5 min, separated by 10–20% pre-cast polyacrylamide gels (Biorad) and transferred to a nitrocellulose membrane (Biorad). Membrane was blocked by 5% milk in Tris-buffered saline (pH 7.4) containing Tween 0.05% (TBST), incubated with primary peptide antibody for 2 h, washed with TBST, incubated for 1 h with horseradish peroxidase labelled anti-rabbit antibody (Jackson Lab) and washed with TBST; peroxidase was detected by SuperSignal West Pico Chemiluminescent Substrate (Pierce).

Sheath preparations. Overnight culture was diluted 1:200 into 200 ml of fresh LB and then shaken at 37 °C for 2.5–3.0 h to reach D_{600} = 1.0–1.5. Cells were cooled on ice, centrifuged for 10 min at 7000g and lysed in 12 ml lysis buffer (150 mM NaCl, 50 mM Tris, pH 7.4, lysozyme 200 μ g ml⁻¹, DNase I 50 μ g ml⁻¹, 1 mM phenylmethylsulphonyl fluoride, 0.5 \times CelLytic B (Sigma), 1% Triton X-100). Cell lysis was complete after incubation for 5–10 min at 37 °C. After cell lysis, samples were cooled on ice and intact cells and cell debris were removed by centrifugation for 15 min at 15,000g. Cleared lysates were subjected to ultraspeed centrifugation at 150,000g for 1 h at 4 °C. Pellets were re-suspended in 0.5 ml of 50 mM Tris, 150 mM NaCl, pH 7.4, supplemented with protease inhibitor cocktail Complete Mini (Roche) and stored at 4 °C or –20 °C for electron microscopy analysis.

Preparation of sheath for mass spectrometry analysis. Sheath for mass spectrometry analysis was prepared from a *flgG* in-frame deletion mutant of the parental 2740-80 strain. Cells were prepared and lysed as described above. To separate the sheath from soluble proteins, the pellet obtained by ultracentrifugation was re-suspended in 12 ml of TN buffer (50 mM Tris, 150 mM NaCl, pH 7.4) and insoluble material removed by a 2 min 15,000g centrifugation step. The sheath was then collected by sequential ultracentrifugation at 150,000g for 1 h. The sheath pellet was again re-suspended in 12 ml TN buffer and subjected to another ultracentrifugation step. After three successive ultracentrifugations, samples typically showed only two major bands on a 10–20% SDS–PAGE. The two detectable bands (20 and 50 kDa), and the areas above and below the bands, were cut out from the gel and analysed by tandem mass spectrometry for peptide identity (Taplin Biological Mass Spectrometry Facility, Harvard).

Peptide-specific antibodies. Antigen-purified rabbit polyclonal antibodies raised against an Hcp peptide (QSGQPSGQRVHKPF) and VipA peptide (MSKEGSAVAPKERIN) were obtained commercially (GenScript). Specificity of the antibodies was tested on *V. cholerae* V52 strains expressing or lacking Hcp protein, or *V. cholerae* 2740-80 strains expressing or lacking VipA.

Fluorescence microscopy. Overnight cultures of *V. cholerae* 2740-80 or V52 strains carrying plasmid pBAD24-VipA-sfGFP or pBAD24-VipA-mCherry2 were diluted 1:100 into fresh LB supplemented with carbenicillin and arabinose (concentrations 0.01%, 0.003% or as indicated) and cultivated for 2.5–3 h to an attenuance of about 1.0. Cells from 100 μ l of the culture were re-suspended in 5 μ l phosphate buffered saline (PBS), spotted on a thin pad of 1% agarose in PBS, covered with a cover slip and immediately imaged at room temperature.

Fluorescence and phase contrast micrographs were captured using a Nikon TE2000 inverted microscope outfitted with a Nikon Intensilight illuminator, a Coolsnap HQ2 charge-coupled device camera from Photometrics and a CFI Plan Apo DM 100 objective lens (1.4 numerical aperture). The sfGFP images were taken by using the ET-GFP filter set (Chroma 49002). The mCherry2 images were taken by using the ET-mCherry filter set (Chroma 49008). Images were captured using Nikon Elements software. Images were collected every 6 or 10 s, using an exposure time of 100–600 ms for fluorescence and about 10–20 ms for phase contrast. Phase contrast imaging was used to refocus automatically between individual time points. Contrast on images for phase and fluorescence channels was adjusted identically for compared image sets and merged using ImageJ 1.45 software (<http://rsb.info.nih.gov/ij/>). Small movement of whole field in time was corrected by registering individual frames using StackReg plugin for ImageJ ('Rigid Body' transformation). The pixel-size was 60 nm.

High-frame-rate fluorescent images were collected with a Nikon Ti-E inverted motorized microscope equipped with a $\times 100$ Plan Apo NA 1.4 objective lens and the Perfect Focus System for continuous maintenance of focus. VipA-sfGFP fluorescence was excited using a Prior Lumen200Pro metal halide epi-fluorescence light source, selected with an ET490/20 \times filter (Chroma) and collected with an ET535/30m filter (Chroma). Two different cameras and acquisition settings were used to collect images. A Hamamatsu ORCA-R2 cooled CCD camera was used to acquire images every 118 ms (exposure time 50 ms, with continuous illumination). A Hamamatsu ORCA-Flash2.8 cooled CMOS camera was used to acquire images every 20 ms (no analogue gain) or 5 ms (8 \times on-chip analogue gain) under continuous illumination light. Both cameras were controlled with Molecular Devices MetaMorph version 7.7 software. Contrast was adjusted identically for compared image sets. All image processing and analyses were done using ImageJ 1.45 software. The pixel size was 67 nm for the ORCA-R2 and 78 nm for the ORCA-Flash2.8 camera.

Plunge-freezing. For ECT, *V. cholerae* 2740-80 wild-type and mutant strains were grown aerobically at 37 °C in LB medium. A 5 ml overnight-culture was diluted 1000-fold and grown to D_{600} = 1.5–2.2. Copper/rhodium electron microscopy grids (R2/2, Quantifoil) were glow-discharged for 1 min. A 20 \times -concentrated bovine serum albumin-treated solution of 10 nm colloidal gold (Sigma) was added to the sample (1:4 v/v) immediately before plunge freezing. A 4 μ l droplet of the mixture was applied to the electron microscopy grid, then automatically blotted and plunge-frozen into a liquid ethane–propane mixture²⁸ using a Vitrobot (FEI Company)³⁶. The grids were stored in liquid nitrogen.

Negative stain electron microscopy. Samples were incubated on carbon-coated grids for about 1 min. Grids were washed in water and stained by 1% uranyl formate. The grids were examined in a JEOL 1200EX transmission electron microscope and images were recorded with an AMT 2k CCD camera.

Electron cryotomography. Tilt series were collected using a Polara 300 kV FEG transmission electron microscope (FEI Company) equipped with an energy filter (slit width 20 eV; Gatan) on a lens-coupled 4k \times 4k UltraCam (Gatan). Pixels on the CCD represented 0.95 nm ($\times 22,500$) or 0.63 nm ($\times 34,000$) at the specimen level. Typically, tilt series were recorded from –60° to +60° with an increment of 1° at 10 μ m under-focus. The cumulative dose of a tilt-series was 180–220 electrons \AA^{-2} (for whole cells) or 80–100 electrons \AA^{-2} (for sheath preparations). Legion²⁹ or UCSF Tomo³⁰ was used for automatic tilt-series acquisition. Three-dimensional reconstructions were calculated using the IMOD software package³¹ or Raptor³².

Sub-tomogram averaging. IMOD³¹ was used to model the centre of the sheath. The program addModPts was run to fill in model points every 8 nm along the tube axis. The PEET software package³³ was used to align and average repeating sub-volumes. Isosurface rendering of the sub-volume averages was done with Chimera³⁴.

35. Metcalf, W. W. *et al.* Conditionally replicative and conjugative plasmids carrying *lacZ* alpha for cloning, mutagenesis, and allele replacement in bacteria. *Plasmid* **35**, 1–13 (1996).
36. Iancu, C. V. *et al.* Electron cryotomography sample preparation using the Vitrobot. *Nature Protocols* **1**, 2813–2819 (2006).

An epigenetic blockade of cognitive functions in the neurodegenerating brain

Johannes Gräff^{1,2,3}, Damien Rei^{1,2}, Ji-Song Guan^{1,2,3}, Wen-Yuan Wang^{1,2,3}, Jinsoo Seo^{1,2}, Krista M. Hennig^{3,4}, Thomas J. F. Nieland³, Daniel M. Fass^{3,4}, Patricia F. Kao⁵, Martin Kahn¹, Susan C. Su^{1,2}, Alireza Samiei¹, Nadine Joseph^{1,2,3}, Stephen J. Haggarty^{3,4}, Ivana Delalle⁵ & Li-Huei Tsai^{1,2,3}

Cognitive decline is a debilitating feature of most neurodegenerative diseases of the central nervous system, including Alzheimer's disease¹. The causes leading to such impairment are only poorly understood and effective treatments are slow to emerge². Here we show that cognitive capacities in the neurodegenerating brain are constrained by an epigenetic blockade of gene transcription that is potentially reversible. This blockade is mediated by histone deacetylase 2, which is increased by Alzheimer's-disease-related neurotoxic insults *in vitro*, in two mouse models of neurodegeneration and in patients with Alzheimer's disease. Histone deacetylase 2 associates with and reduces the histone acetylation of genes important for learning and memory, which show a concomitant decrease in expression. Importantly, reversing the build-up of histone deacetylase 2 by short-hairpin-RNA-mediated knockdown unlocks the repression of these genes, reinstates structural and synaptic plasticity, and abolishes neurodegeneration-associated memory impairments. These findings advocate for the development of selective inhibitors of histone deacetylase 2 and suggest that cognitive capacities following neurodegeneration are not entirely lost, but merely impaired by this epigenetic blockade.

Long-lasting forms of memories require stable gene expression changes³, which are in part orchestrated by chromatin-templated epigenetic processes⁴. Of the epigenetic modifications identified so far in the nervous system, histone acetylation has been unequivocally associated with facilitating learning and memory⁴. Acetylation diminishes the electrostatic affinity between neighbouring histones and the DNA and, consequently, can promote a more open chromatin structure that allows for memory-related gene transcription⁵.

Over the past decade, several studies have reported sporadic cases of reduced histone acetylation in animal models of neurodegeneration that are characterized by cognitive decline, including models of Alzheimer's disease⁴. Accordingly, pharmacological treatments aimed at increasing histone acetylation have shown promising results in reversing cognitive deficits in some of these models, predominantly by the use of non-selective histone deacetylase (HDAC) inhibitors⁶. However, the causative agent of such memory-impairing histone acetylation changes, and, hence, the best targets for pharmacological strategies, remain unknown. One likely candidate is HDAC2, a class I HDAC that negatively regulates memory and synaptic plasticity in the healthy mouse brain^{7,8}.

To investigate whether HDAC2 mediates cognitive deficits associated with neurodegeneration, we measured its abundance in CK-p25 mice^{9,10}, which inducibly and forebrain-specifically overexpress p25, a truncated version of p35. p25 aberrantly activates cyclin-dependent kinase 5 (CDK5), and is implicated in various neurodegenerative diseases¹¹, including Alzheimer's disease¹². After 6 weeks of p25 induction, CK-p25 mice display Alzheimer's-disease-related pathologies such as neuronal loss⁹, β -amyloid accumulation¹⁰, reactive astrogliosis⁹ and reduced

synaptic density¹³, most prominently in the hippocampus and the cortex, two brain areas important for memory formation and storage, respectively¹⁴. Accordingly, 6-week-induced CK-p25 (hereafter referred to as CK-p25) mice also display spatial and associative memory deficits¹³.

Using immunohistochemistry and western blot analysis, we found that HDAC2 was significantly increased in neuronal nuclei in hippocampal area CA1 in CK-p25 mice compared with control littermates (Fig. 1a, d, e; see Supplementary Fig. 2a for a specificity control for the HDAC2 signal). No changes in HDAC2 were observed in hippocampal area CA3 or the dentate gyrus (Supplementary Fig. 3a, b), explaining the overall marginal increase in the entire hippocampus. Interestingly, this effect appears to be non-cell-autonomous, as both p25-positive and p25-negative cells displayed elevated HDAC2 (Fig. 1a). In contrast, levels of the structurally highly related HDAC1, and of HDAC3, another class I HDAC involved in memory formation¹⁵, were not changed (Fig. 1b, c, e). Furthermore, HDAC2, but not HDAC1 or HDAC3, was also increased in the prefrontal cortex of CK-p25 mice (Supplementary Fig. 4), whereas in the amygdala, a brain area not affected by neurodegeneration in the CK-p25 mice, its levels remained unchanged (Supplementary Fig. 3c). This neurodegeneration-associated increase of HDAC2 was confirmed in another mouse model of Alzheimer's-disease-related pathologies and cognitive decline, the 5XFAD mouse^{16,17} (Supplementary Fig. 5).

Next, we aimed to determine the functional consequences of elevated HDAC2. Because HDAC2 has been shown to associate with the promoter region of genes involved in memory and synaptic plasticity⁷, we proposed that it is enriched at these genes following neurodegeneration. Of the known HDAC2 targets⁷, we focused on those that, in several independent studies, had been demonstrated to be downregulated in the human brain with Alzheimer's disease (Supplementary Table 1). These include the immediate-early genes *Arc*, *Bdnf* exons I, II and IV, *Egr1*, *Homer1*, and *Cdk5*, implicated in learning and memory, and genes related to synaptic plasticity such as the glutamate receptor subunits *GluR1*, *GluR2*, *NR2A* and *NR2B* (also known as *Gria1*, *Gria2*, *Grin2a* and *Grin2b*), as well as *Nfl* (neurofilament light chain, also known as *Nefl*), *Syp* (synaptophysin) and *Syt1* (synaptotagmin 1). Using chromatin immunoprecipitation (ChIP, for primers see Supplementary Table 2), we found that HDAC2 is significantly enriched at these genes in the CK-p25 hippocampus, the exception being the promoter regions of the activity-dependent *Bdnf* exons I and II¹⁸, and the housekeeping genes β -actin, β -globin and β -tubulin (Fig. 1f). In contrast, binding of HDAC1 and HDAC3 was unaltered (Supplementary Fig. 6a, b). Interestingly, in agreement with previous reports showing that HDAC2 can also bind to a gene's coding region¹⁹, we also found HDAC2 more abundantly bound to the coding sequence of the same genes (see Supplementary Fig. 7 and Supplementary Table 3 for primer sequences).

¹Picower Institute for Learning and Memory, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ²Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ³Stanley Center for Psychiatric Research, Broad Institute of Harvard University and Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. ⁴Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA. ⁵Department of Pathology and Laboratory Medicine, Boston University School of Medicine, Boston, Massachusetts 02118, USA.

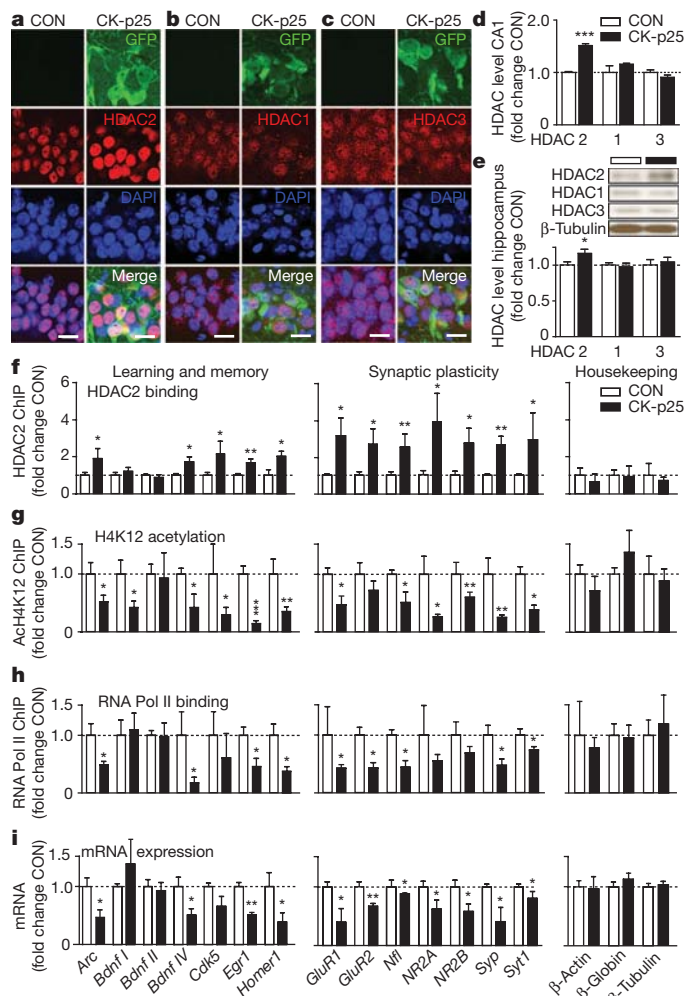


Figure 1 | Elevated HDAC2 levels epigenetically block the expression of neuroplasticity genes during neurodegeneration. **a–c**, Representative immunohistochemical images depicting HDAC1–3 levels in area CA1 of CK-p25 mice and control littermates; scale bar, 20 μ m. **d**, Quantitative assessment of **a–c** ($n = 3–6$ slices from three or four mice each). **e**, Representative western blot images and quantification of HDAC1–3 in the CK-p25 and control hippocampus ($n = 6–9$ mice each). **f–h**, Quantitative PCR results of (f) HDAC2-, (g) ACh4K12- and (h) RNA Pol II-immunoprecipitated chromatin at the promoter of neuroplasticity and housekeeping genes in the CK-p25 and control hippocampus. **i**, Quantitative RT-PCR results of the same genes (**f–i**, $n = 4–8$ animals each). * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; values are mean \pm s.e.m.

We next assessed the acetylation of several histone (H) residues in the promoter region of these genes, for which acetylation has been shown to be important for learning, memory, and synaptic plasticity, such as H2B lysine (K) 5, H3K14, H4K5 and H4K12 (ref. 4). ChIP analyses revealed a hypoacetylation for all residues at the neuroplasticity genes (Fig. 1g and Supplementary Fig. 6c–e), albeit to different extents. Importantly, the acetylation of housekeeping genes was not altered. The effects of elevated HDAC2 levels further appear to be restricted to histones, as we found no overall acetylation changes on other proteins regulated by this modification, such as tau (also known as MAPT), protein 53 (p53; also known as TP53) and tubulin, nor in overall nuclear or cytoplasmic protein acetylation (Supplementary Fig. 8).

Next, to determine the functional consequences of promoter hypoacetylation, we assessed the binding of activated (that is, phosphorylated) RNA polymerase II (RNA Pol II), and found it to be markedly reduced (Fig. 1h). This prompted us to measure the messenger RNA (mRNA) expression of these genes by quantitative reverse-transcription PCR (RT-PCR) (primers in Supplementary Table 4). We found reduced expression for all genes with elevated HDAC2 binding and

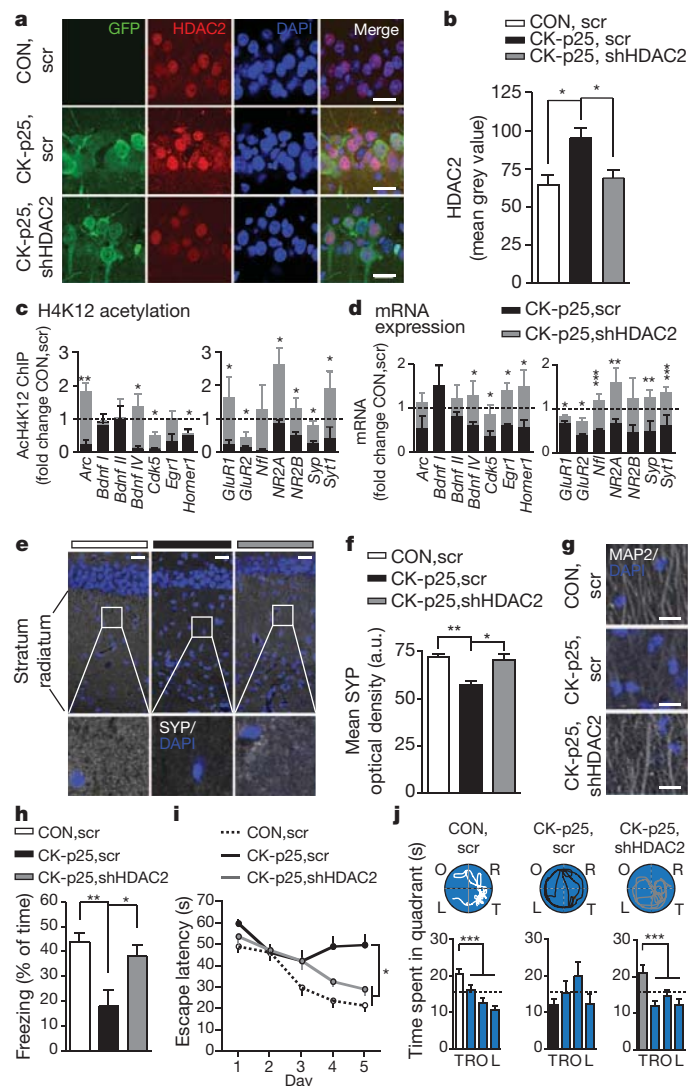


Figure 2 | Reducing HDAC2 levels alleviates memory deficits.

a, Representative immunohistochemical images depicting HDAC2 in hippocampal area CA1 of CK-p25, shHDAC2, CK-p25, scr and CON, scr animals; scale bar, 20 μ m. **b**, Quantitative assessment of **a**, $n = 4$ or 5 sections from four mice each. **c**, Quantitative PCR results of ACh4K12-immunoprecipitated chromatin in CK-p25, scr and CK-p25, shHDAC2 compared with CON, scr mice. **d**, Quantitative RT-PCR results of the same genes. (**c**, **d**, $n = 4–6$ animals each). **e**, **g**, Representative immunohistochemical images depicting (e) SYP and (g) MAP2 immunoreactivity in the hippocampus stratum radiatum; scale bars: **e**, 25 μ m; **g**, 20 μ m. **f**, Quantitative assessment of **e**, $n = 4$ mice each; a.u., arbitrary units. **h**, Freezing responses of CON, scr ($n = 18$), CK-p25, scr ($n = 16$) and CK-p25, shHDAC2 ($n = 16$) mice 24 h after contextual fear conditioning. **i**, Escape latencies in a water maze task of CON, scr ($n = 19$), CK-p25, scr ($n = 17$) and CK-p25, shHDAC2 ($n = 19$) animals. Data points are averages of two trials per day. **j**, Representative swim traces and time spent per quadrant during the water maze test (T, target quadrant; R, right; O, opposite; L, left of target). * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; values are mean \pm s.e.m.

a concomitant decrease in histone acetylation and RNA Pol II binding (Fig. 1i). Of note, HDAC2 probably acts with the transcriptional repressor complexes CoREST, NuRD and SIN3, as we found increased association of HDAC2 with these complexes in hippocampal CK-p25 extracts by co-immunoprecipitation (Supplementary Fig. 9). Taken together, these results indicate that HDAC2 mediates a local chromatin compaction of neuroplasticity genes, which decreases their expression and may contribute to cognitive decline during neurodegeneration.

To examine causally such a possibility, we generated adeno-associated viral vectors carrying either short-hairpin RNAs (shRNAs) directed against HDAC2 or scrambled control shRNAs (Supplementary Fig.

10a, b). The knockdown efficiency of the shRNAs in culture was about 25–30% (Supplementary Fig. 10c–f), ideally suited for the targeted normalization of the 20–50% increase of HDAC2 in the CK-p25 mice. We injected these vectors bilaterally into hippocampal area CA1 of 2-week-induced CK-p25 and control mice, and assessed HDAC2 levels 4 weeks after viral injection (Supplementary Fig. 10g). CK-p25 animals injected with an shRNA against HDAC2 (CK-p25, shHDAC2) showed reduced HDAC2 levels compared with CK-p25 animals injected with control scrambled shRNA (CK-p25, scr), indistinguishable from control mice injected with scrambled shRNA (CON, scr) (Fig. 2a, b). Protein levels of HDAC1 and HDAC3 remained unchanged (Supplementary Fig. 11). We observed transduction efficiencies of 53–61% (mean \pm s.e.m. 57.4 ± 2.5 ; $n = 3$ –4 mice per group analysed) and comparable infection rates in both control and CK-p25 hippocampi (Supplementary Fig. 10h).

Next, we sought to determine whether reducing HDAC2 levels would alter the promoter histone acetylation and mRNA expression of neuroplasticity genes. We found that H4K12 acetylation was significantly enhanced on most of these genes, and they showed increased expression (Fig. 2c, d). Importantly, most of these genes showed comparable, or even higher, expression in CK-p25, shHDAC2 mice compared with CON, scr animals (black dotted lines).

Based on these findings, we investigated whether such regained chromatin and transcriptional plasticity might translate into morphological

and physiological changes in the surviving neurons. To this end, we measured synaptic density by SYP immunohistochemistry, labelling the presynaptic terminals of functional synapses, and dendritic abundance by microtubule-associated protein 2 (MAP2) immunoreactivity. We found that, whereas SYP and MAP2 levels were reduced in the stratum radiatum of CK-p25, scr animals (as previously described for CK-p25 mice^{13,20}), their abundance was markedly increased in CK-p25, shHDAC2 animals, to levels comparable to CON, scr animals (Fig. 2e–g). Interestingly, however, the number of surviving neurons, as evidenced by NeuN immunohistochemistry, was not altered by HDAC2 reduction (Supplementary Fig. 12). Next, we measured synaptic plasticity by electrophysiological recordings and observed that long-term potentiation in CA1 neurons was undistinguishable between CK-p25, shHDAC2 and CON, scr animals, but significantly improved over CK-p25, scr animals. A similar effect was observed for basal synaptic plasticity (Supplementary Fig. 13). Together, these data indicate that, although HDAC2 normalization did not impact neuronal survival, it did reinstate morphological and synaptic plasticity in the surviving neurons.

We suggested that the reduction of HDAC2 would also counteract the cognitive deficits associated with neurodegeneration. For this, we assessed associative and spatial memory on a fear-conditioning and water maze task, respectively, two types of hippocampus-dependent memory that are severely perturbed in CK-p25 animals¹³. We observed

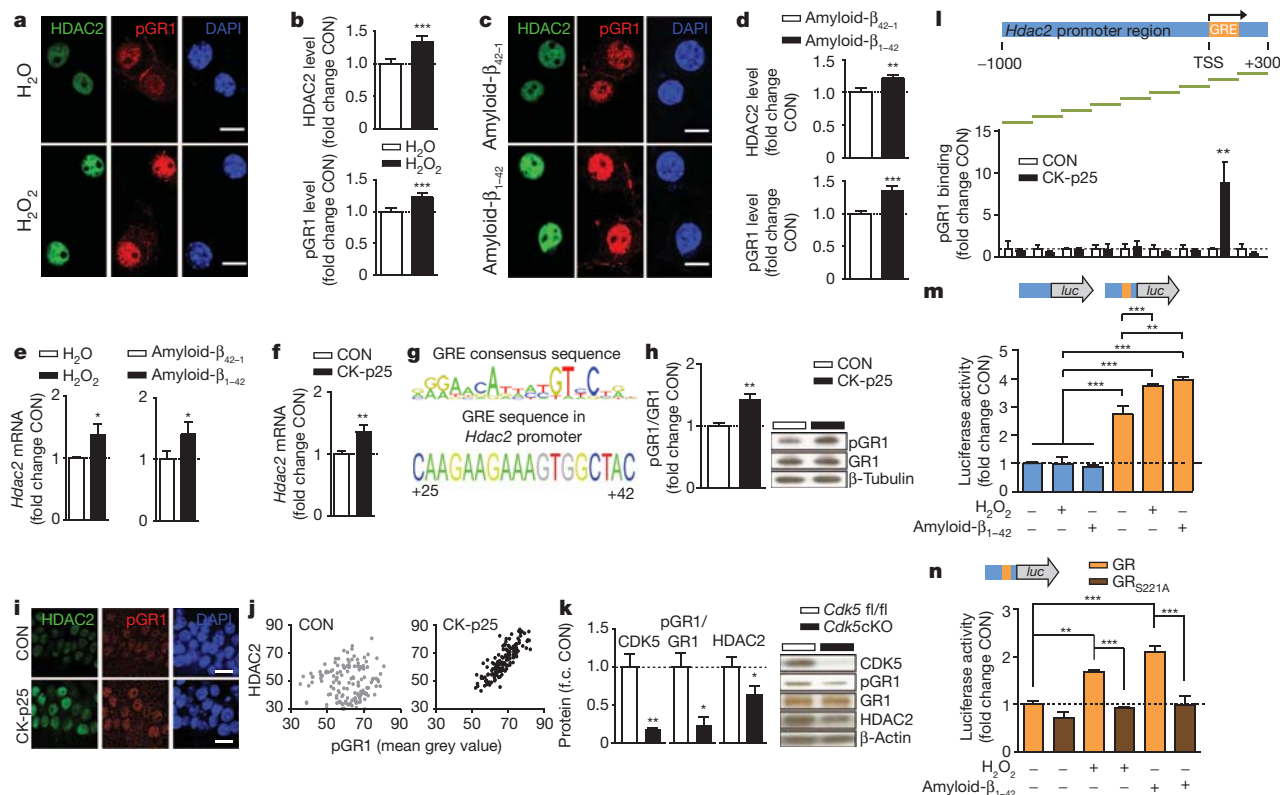


Figure 3 | Neurotoxic insults increase HDAC2 through stress elements in its promoter. **a, c,** Representative pictures of HDAC2 and pGR1 labelling of primary hippocampal neurons treated with **(a)** H_2O_2 and **(c)** amyloid- β -oligomers ($n = 20$ –40 neurons per group); scale bar, 10 μ m. **b, d,** Quantification of **a** and **c**. **e, f,** Quantitative RT-PCR results showing *Hdac2* expression in **(e)** H_2O_2 - and amyloid- β -treated primary hippocampal neurons, and **(f)** in the CK-p25 hippocampus ($n = 7$ –9 mice each). **g,** Alignment of the vertebrate GRE consensus sequence with the GRE in the proximal promoter of mouse *Hdac2*. **h,** Quantification and representative western blot images of hippocampal extracts of CK-p25 versus control mice ($n = 3$ each). **i,** Representative images of immunohistochemical labelling of pGR1 and HDAC2 in the CK-p25 hippocampus ($n = 3$ –6 slices from three mice each); scale bar, 20 μ m. **j,** Regression analysis of **i** showing a significant correlation between pGR1 and HDAC2 in CK-p25 ($R^2 = 0.686$, $P \leq 0.001$), but

not control mice ($R^2 = 0.019$, not significant). **k,** Quantification and representative western blot images of *Cdk5* KO and control *Cdk5* fl/fl forebrain extracts ($n = 3$ each). f.c., fold change. **l,** Quantitative PCR results of pGR1-immunoprecipitated chromatin around the GRE in a 1.3-kb-wide *Hdac2* promoter region (schematically shown above the graph; TSS, transcriptional start site) in the CK-p25 and control hippocampus ($n = 3$ –6 animals each); green lines represent fragments amplified by primer pairs. **m,** Luciferase activity of CAD cells transfected with the *Hdac2* promoter with (orange) or without (blue) GRE (schematic of constructs shown above graph), and treated with H_2O_2 and amyloid- β_{1-42} . **n,** Luciferase activity of CAD cells transfected with *Hdac2*-GRE in the presence of endogenous glucocorticoid receptor or of cotransfected GR_{S211A}. *In vitro* results are from at least three independent experiments. * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; values are mean \pm s.e.m.

that associative memory of CK-p25, shHDAC2 animals returned to levels of CON, scr animals (Fig. 2h). Likewise, CK-p25, shHDAC2 animals showed significantly reduced escape latencies compared with CK-p25, scr animals during training in the water maze (Fig. 2i) and, 24 h later, they spent significantly more time in the target quadrant, indistinguishable from the performance of CON, scr mice (Fig. 2j). Overall, swimming behaviour was similar between the different groups (Supplementary Fig. 14a, b), and altering HDAC2 levels per se did not affect locomotor activity or anxiety as assessed by an open field test (Supplementary Fig. 14c–f). Together, these results indicate that elevated HDAC2 levels are causally involved in the cognitive decline associated with neurodegeneration in CK-p25 mice, but that the prevention of HDAC2 upregulation rescues memory capacities.

To gain insight into the mechanisms underlying the increase in HDAC2, we exposed primary hippocampal neurons to neurotoxic stimuli characteristic of Alzheimer's-disease-related neurodegeneration, hydrogen peroxide (H_2O_2) and amyloid- β oligomers^{21,22}. As revealed by immunocytochemistry and western blot analysis, treatment with either H_2O_2 or amyloid- β_{1-42} , but not control amyloid- β_{42-1} , oligomers was sufficient to increase HDAC2 (Fig. 3a–d, left panels, and Supplementary Fig. 15a–d). Importantly, both neurotoxic stimuli increased *Hdac2* at the mRNA level (Fig. 3e), and increased *Hdac2* transcription was also evident in the CK-p25 hippocampus (Fig. 3f), suggesting the involvement of transcriptional mechanisms.

This prompted us to screen the *Hdac2* promoter for potential binding sites of transcriptional regulators. Using transcription factor binding databases²³, we found a well-conserved recognition element for the glucocorticoid receptor 1 (GR1, also known as NR3C1) in the proximal promoter region of *Hdac2* (Fig. 3g). Glucocorticoid receptors are activated by phosphorylation following behavioural or cellular stress and, upon binding to the glucocorticoid responsive element (GRE) in a gene's promoter region, they can act as transcriptional activators or repressors, depending, in part, on the residue phosphorylated²⁴. Of its known phosphorylation sites, serine (S) 211 has been robustly associated with activated forms of GR1 (ref. 25).

Based on this knowledge, we examined whether S211 phosphorylation on GR1 was increased after neurotoxic insults *in vitro*, and in the CK-p25 brain *in vivo*. Immunocytochemical labelling and western blot analysis of cultured hippocampal neurons following H_2O_2 and amyloid- β_{1-42} treatment revealed a significantly increased phosphorylation of GR1 (pGR1) on S211 compared with control conditions (Fig. 3a–d, middle panels, and Supplementary Fig. 15a–d; see Supplementary Fig. 2b for anti-pGR1_{S211} specificity). Furthermore, the CK-p25 hippocampus showed similarly increased pGR1 levels (Fig. 3h, i and Supplementary Fig. 16). Remarkably, we observed that the increase in pGR1 occurred concurrently with that of HDAC2 following neurotoxicity (Fig. 3a, c, i, j) and, using forebrain extracts of conditional *Cdk5* knockout (*Cdk5cKO*) and control mice²⁶, we identified CDK5 as a GR1 kinase *in vivo* (Fig. 3k).

We then sought to determine whether GR1 phosphorylation increases *Hdac2* transcription. We first examined the binding of pGR1 to the *Hdac2* promoter by pGR1-ChIP and primer pairs (Supplementary Table 2) spanning a region from –1000 to +300 base pairs around the *Hdac2* transcriptional start site (Fig. 3l, top). We found that pGR1 binding to the *Hdac2*-GRE was significantly increased in the CK-p25 hippocampus (Fig. 3l, bottom). Similar results were obtained after amyloid- β_{1-42} treatment of primary hippocampal cultures (Supplementary Fig. 15e). Second, to determine whether GR1 is directly capable of transcriptionally activating *Hdac2*, we cloned the *Hdac2* promoter with and without the GRE into a luciferase construct and tested its activity in CAD cells, a primary neuron-like cell line. We found that the presence of the GRE alone increased the luciferase activity by approximately threefold but that, upon the addition of a constitutively active form of glucocorticoid receptor, GR526 (see Methods), this activity was further doubled. However, without the *Hdac2*-GRE, the addition of GR526 had no effect (Supplementary Fig. 17). Importantly, we found

that the *Hdac2*-GRE also responds to neurotoxic stimuli. When CAD cells were treated with either H_2O_2 or amyloid- β_{1-42} , luciferase activity was significantly increased, but only with the GRE present (Fig. 3m, orange bars). Furthermore, the capability of GR to activate *Hdac2* critically depends upon its phosphorylation. When S211 was mutated to alanine (GR_{S211A}), the glucocorticoid receptor was no longer capable of activating *Hdac2* *in vitro* (Fig. 3n). This dependency on S211 phosphorylation also occurs *in vivo*, as *Cdk5cKO* forebrain extracts²⁶ had reduced pGR1 and HDAC2 levels (Fig. 3k). Taken together, these results indicate that Alzheimer's-disease-related neurotoxic stimuli lead to an increase in *Hdac2* gene transcription by mechanisms involving glucocorticoid receptor activation and its interaction with the *Hdac2*-GRE.

Lastly, to assess the relevance of these findings in humans, we compared HDAC2 abundance in post-mortem brain samples from patients with varying degrees of non-familial Alzheimer's disease, the most common form of neurodegeneration-associated dementia worldwide²⁷. The cases used here (Supplementary Table 5) are defined by the Braak and Braak stages²⁸, which are characterized by the accumulation of hyperphosphorylated tau protein in the cortices (Supplementary Fig. 18), and by increasing neurodegeneration²⁸ and cognitive impairment²⁹. We found that, in all Alzheimer's-disease-related Braak and Braak stages, HDAC2 levels were significantly elevated in hippocampal area CA1 (Fig. 4a, d) and the entorhinal cortex (Supplementary Fig. 19a, d), which are among the earliest and most affected brain areas in Alzheimer's disease² and crucial for memory formation and storage¹⁴. HDAC2 accumulation was visible beginning

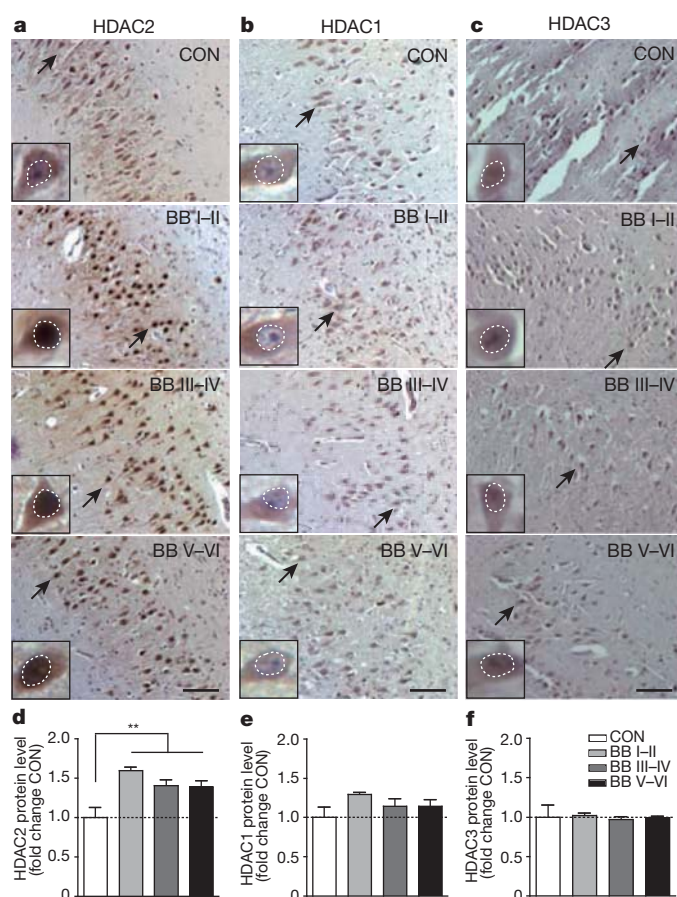


Figure 4 | HDAC2 expression is increased in patients with Alzheimer's disease. a–c, Representative immunohistochemical images depicting nuclear HDAC1–3 levels (white dotted circles) in neurons (arrow points to magnified neuron in inset) of hippocampal area CA1 from patients with Braak and Braak (BB) stages I–II ($n = 4$), III–IV ($n = 7$) and V–VI ($n = 8$) compared with healthy BB0 control brains (CON, $n = 7$); scale bar, 100 μ m. d–f, Quantitative assessment of a–c. $^{**}P \leq 0.01$; values are mean \pm s.e.m.

at Braak and Braak stage I/II, implicating it as an early event in the progression of Alzheimer's disease. In contrast, levels of HDAC1 and HDAC3 were not altered (Fig. 4b, c, e, f and Supplementary Fig. 19b, c, e, f). Thus, elevated levels of HDAC2 may also accompany the cognitive decline of the human neurodegenerating brain.

The findings presented in this study describe that epigenetic mechanisms substantially contribute to the cognitive decline associated with Alzheimer's disease-related neurodegeneration. Although it is well documented that neuronal loss and amyloid- β - or tau-induced neurotoxicity acutely disable synaptic functions, in turn leading to cognitive deficits^{1,2,30}, the HDAC2-mediated epigenetic blockade of neuroplasticity-related gene expression could delineate a process by which memory functions become permanently impaired in the Alzheimer's disease brain (Supplementary Fig. 1a). This blockade appears to be induced by GR1; thus, glucocorticoid receptors may function as molecular mediators between neurodegeneration-associated neurotoxic stressors and cognitive impairment (Supplementary Fig. 1b).

Intriguingly, our findings may also provide a potential explanation, at least in part, as to why, in some clinical trials, cognitive impairments in patients with Alzheimer's disease persist despite successful amyloid- β clearance²: once the epigenetic blockade is in place, reducing amyloid- β generation and deposition alone may not be sufficient to rescue against cognitive dysfunction. A more efficacious strategy may therefore lie in the combination of amyloid- β reduction with the inhibition of HDAC2. By extension, these findings pinpoint HDAC2 as the probable target of non-selective HDAC inhibitors that counteract cognitive decline in mouse models of Alzheimer's disease⁶ and, as a result, strongly advocate for the development of HDAC2-selective inhibitors. Finally, our finding that HDAC2 inhibition probably reinstates transcriptional, morphological and synaptic plasticity in the surviving neurons of the neurodegenerating brain raises hope that such plasticity is not irrevocably lost, but merely constrained by the epigenetic blockade.

METHODS SUMMARY

Please refer to Methods for more detail. Human material was used with informed consent from all donors, and all mouse work was approved by the Committee for Animal Care of the Division of Comparative Medicine at the Massachusetts Institute of Technology.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 July 2011; accepted 13 January 2012.

Published online 29 February 2012.

- Walsh, D. M. & Selkoe, D. J. Deciphering the molecular basis of memory failure in Alzheimer's disease. *Neuron* **44**, 181–193 (2004).
- Holtzman, D. M., Morris, J. C. & Goate, A. M. Alzheimer's disease: the challenge of the second century. *Sci. Transl. Med.* **3**, 77sr1 (2011).
- Kandel, E. R. The molecular biology of memory storage: a dialogue between genes and synapses. *Science* **294**, 1030–1038 (2001).
- Gräff, J., Kim, D., Dobbin, M. M. & Tsai, L. H. Epigenetic regulation of gene expression in physiological and pathological brain processes. *Physiol. Rev.* **91**, 603–649 (2011).
- Brownell, J. E. & Allis, C. D. Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. *Curr. Opin. Genet. Dev.* **6**, 176–184 (1996).
- Kazantsev, A. G. & Thompson, L. M. Therapeutic application of histone deacetylase inhibitors for central nervous system disorders. *Nature Rev. Drug Discov.* **7**, 854–868 (2008).
- Guan, J. S. *et al.* HDAC2 negatively regulates memory formation and synaptic plasticity. *Nature* **459**, 55–60 (2009).
- Akhtar, M. W. *et al.* Histone deacetylases 1 and 2 form a developmental switch that controls excitatory synapse maturation and function. *J. Neurosci.* **29**, 8288–8297 (2009).
- Cruz, J. C., Tseng, H. C., Goldman, J. A., Shih, H. & Tsai, L. H. Aberrant Cdk5 activation by p25 triggers pathological events leading to neurodegeneration and neurofibrillary tangles. *Neuron* **40**, 471–483 (2003).
- Cruz, J. C. *et al.* p25/cyclin-dependent kinase 5 induces production and intraneuronal accumulation of amyloid beta *in vivo*. *J. Neurosci.* **26**, 10536–10541 (2006).
- Cruz, J. C. & Tsai, L. H. A Jekyll and Hyde kinase: roles for Cdk5 in brain development and disease. *Curr. Opin. Neurobiol.* **14**, 390–394 (2004).

- Patrick, G. N. *et al.* Conversion of p35 to p25 deregulates Cdk5 activity and promotes neurodegeneration. *Nature* **402**, 615–622 (1999).
- Fischer, A., Sananbenesi, F., Pang, P. T., Lu, B. & Tsai, L. H. Opposing roles of transient and prolonged expression of p25 in synaptic plasticity and hippocampus-dependent memory. *Neuron* **48**, 825–838 (2005).
- Frankland, P. W. & Bontempi, B. The organization of recent and remote memories. *Nature Rev. Neurosci.* **6**, 119–130 (2005).
- McQuown, S. C. *et al.* HDAC3 is a critical negative regulator of long-term memory formation. *J. Neurosci.* **31**, 764–774 (2011).
- Oakley, H. *et al.* Intraneuronal beta-amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial Alzheimer's disease mutations: potential factors in amyloid plaque formation. *J. Neurosci.* **26**, 10129–10140 (2006).
- Kimura, R., Devi, L. & Ohno, M. Partial reduction of BACE1 improves synaptic plasticity, recent and remote memories in Alzheimer's disease transgenic mice. *J. Neurochem.* **113**, 248–261 (2010).
- Lauterborn, J. C. *et al.* Differential effects of protein synthesis inhibition on the activity-dependent expression of BDNF transcripts: evidence for immediate-early gene responses from specific promoters. *J. Neurosci.* **16**, 7428–7436 (1996).
- Sun, J. M., Chen, H. Y. & Davie, J. R. Differential distribution of unmodified and phosphorylated histone deacetylase 2 in chromatin. *J. Biol. Chem.* **282**, 33227–33236 (2007).
- Fischer, A., Sananbenesi, F., Wang, X. Y., Dobbin, M. & Tsai, L. H. Recovery of learning and memory is associated with chromatin remodelling. *Nature* **447**, 178–182 (2007).
- Markesbery, W. R. & Carney, J. M. Oxidative alterations in Alzheimer's disease. *Brain Pathol.* **9**, 133–146 (1999).
- Yankner, B. A. Mechanisms of neuronal degeneration in Alzheimer's disease. *Neuron* **16**, 921–932 (1996).
- Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W. W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* **32**, D91–D94 (2004).
- Chrousos, G. P. & Kino, T. Glucocorticoid signaling in the cell. Expanding clinical implications to complex human behavioral and somatic disorders. *Ann. NY Acad. Sci.* **1179**, 153–166 (2009).
- Kino, T. *et al.* Cyclin-dependent kinase 5 differentially regulates the transcriptional activity of the glucocorticoid receptor through phosphorylation: clinical implications for the nervous system response to glucocorticoids and stress. *Mol. Endocrinol.* **21**, 1552–1568 (2007).
- Guan, J. S. *et al.* Cdk5 is required for memory function and hippocampal plasticity via the cAMP signaling pathway. *PLoS ONE* **6**, e25735 (2011).
- Citron, M. Alzheimer's disease: strategies for disease modification. *Nature Rev. Drug Discov.* **9**, 387–398 (2010).
- Braak, H. & Braak, E. Diagnostic criteria for neuropathologic assessment of Alzheimer's disease. *Neurobiol. Aging* **18**, S85–S88 (1997).
- Grober, E. *et al.* Memory and mental status correlates of modified Braak staging. *Neurobiol. Aging* **20**, 573–579 (1999).
- Palop, J. J. & Mucke, L. Amyloid-beta-induced neuronal dysfunction in Alzheimer's disease: from synapses toward neural networks. *Nature Neurosci.* **13**, 812–818 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Mungenast, S. Jemielity, R. Madabushi, F. Calderon de Anda and M. Horn for reading the manuscript, A.M. for manuscript editing, M. Eichler for mouse colony maintenance, K. Fitch for sectioning the human brain samples and M.H. for quantification of Fig. 2a. This work was partly supported by the Stanley Medical Research Institution (to S.J.H. and L.-H.T.), National Institutes of Health/National Institute on Drug Abuse (RO1DA028301, to S.J.H.) and National Institutes of Health/National Institute of Neurological Disorders and Stroke (RO1NS078839, to L.-H.T.). J.G. was supported by a Bard Richmond fellowship and by the Swiss National Science Foundation, W.Y.W. by the Simons Foundation and M.K. by the Theodor und Ida Herzog-Egli foundation. L.-H.T. is an investigator of the Howard Hughes Medical Institute.

Author Contributions This study was designed by J.G. and L.-H.T., and directed and coordinated by L.-H.T. J.G. planned and performed the *in vitro*, CK-p25 and Cdk5cKO mouse and human *in vivo* biochemical characterization, and all behavioural experiments. D.R. planned and contributed to the *in vitro* and CK-p25 *in vivo* experiments, generated the GR526 and the shRNA constructs, and contributed to the stereotaxic injections. J.S.G. initiated and contributed to the CK-p25 biochemical characterization, and performed the 5XFAD and HDAC2^{-/-} experiments. W.Y.W. generated the luciferase constructs. J.S. performed the electrophysiological experiments. K.M.H., T.J.F.N., D.F. and S.J.H. characterized the shRNA constructs. M.K. contributed to the quantitative RT-PCR experiments and performed the quantification of the human data. S.C.S. performed the site-directed mutagenesis. A.S. contributed to the immunohistochemistry and the quantitative reverse transcription/quantitative PCR experiments. N.J. contributed to the behavioural and quantitative reverse transcription/quantitative PCR experiments. P.F.K. and I.D. provided the human samples and contributed to the optimization of their staining. The manuscript was written by J.G. and L.-H.T. and commented on by all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.-H.T. (lhtsai@mit.edu).

METHODS

Human material. Human material was obtained from the Massachusetts Alzheimer Disease Research Center at Massachusetts General Hospital (protocol number 2004-P-001613/4) and from the Boston Medical Center (H-24454) with informed consent by all donors. The cases analysed in this study are strictly defined according to the Braak and Braak staging²⁸, with control cases being brains completely devoid of tau pathologies (Braak and Braak stage 0)³⁸. Hippocampal blocks (at the level of the lateral geniculate nucleus) were fixed in paraformaldehyde, paraffin-embedded and sectioned at 10 μ m thickness. After antigen retrieval (Biogenex), brain slices were rehydrated using xylene/ethanol, blocked with 5% milk serum in TBS (pH 7.4) at room temperature and incubated at 4 °C overnight in TBS containing 5% milk-serum in the following antibodies: amyloid- β , tau (Dako), HDAC1, HDAC2 (Abcam) or HDAC3 (Santa Cruz). Next, they were washed in PBS (pH 7.4), incubated in HRP-conjugated secondary antibodies (Biogenex) and visualized with DAB (Biogenex). Slides were counterstained with haematoxylin and dehydrated with ethanol/xylene. Images were quantified by an experimenter blind to Braak and Braak stages using ImageJ 1.42q.

Animal models. All mouse work was approved by the Committee for Animal Care of the Division of Comparative Medicine at the Massachusetts Institute of Technology. Adult (3–6 months old) male double transgenic CK-p25 mice^{9,10,13}, 6-month-old male transgenic 5XFAD mice¹⁶, and 3- to 4-month-old *Cdk5*KO mice²⁶ and their respective control littermates were used for the experiments unless otherwise noted; for CK-p25 mice, all behavioural experiments took place between 6 and 8 weeks of p25 induction, the time when cognitive deficits are first visible¹³. Behavioural experiments were conducted blindly and essentially as described¹³. Open field behaviour was monitored using the VersaMax system (Accuscan) for 20 min. For fear conditioning, mice were put in the conditioning chamber (TSE systems) for 3 min, after which they received a one-time 2 s footshock (0.8 mA). Animals were then left in the box for another 30 s. Twenty-four hours later, the mice were put into the same box and their freezing behaviour was scored during 3 min. For the water maze that took place in a round tank (1.2 m in diameter) filled with white opaque water, mice were first habituated to the task, with the platform being visible for two trials. During habituation and the acquisition phase, mice were allowed to swim for 60 s or until they reached the platform (monitored by HVS Image). Animals that did not reach the platform after 60 s were gently guided towards it; all animals were allowed to remain on the platform for 15 s. For testing, mice were put back into the water without the platform 24 h after that last training session, from a starting position different of the last starting position during the acquisition phase, and their time spent in each quadrant was recorded (HVS Image).

Electrophysiology. To record field excitatory postsynaptic potentials, transverse hippocampal slices were prepared from CON, scr, CK-p25, scr and CK-p25, shHDAC2 mice. In brief, the brain was rapidly removed and transferred to ice-cold, oxygenated (95% O₂ and 5% CO₂) cutting solution containing 211 mM sucrose, 3.3 mM KCl, 1.3 mM NaH₂PO₄, 0.5 mM CaCl₂, 10 mM MgCl₂, 26 mM NaHCO₃ and 11 mM glucose. Hippocampal slices were cut with a VT1000S vibratome (Leica) and transferred for recovery to a holding chamber containing oxygenated artificial cerebrospinal fluid consisting of 124 mM NaCl, 3.3 mM KCl, 1.3 mM NaH₂PO₄, 2.5 mM CaCl₂, 1.5 mM MgCl₂, 26 mM NaHCO₃ and 11 mM glucose at 28–30 °C for at least 1 h before recording. CA1 field potentials evoked by Schaffer collateral stimulation were measured. After recording of a stable baseline (at least 20 min), long-term potentiation was induced by four episodes of theta burst stimulation (TBS) with 10 s intervals. TBS consisted of ten bursts (each with four pulses at 100 Hz) of stimuli delivered every 200 ms. Recordings were performed using an AM-1800 microelectrode amplifier (A-M systems) and a Digidata 1440A analogue to digital converter (Axon Instruments). All data were digitized and analysed by the use of pClamp10 software (Axon Instruments). Basal synaptic input/output relationship was obtained by plotting field excitatory postsynaptic potential slopes against stimulation intensities. All experiments were performed by an experimenter blind to treatment groups.

In vitro studies. Primary mouse hippocampal neuronal cultures (days in vitro 14–17) were treated with sense and antisense amyloid- β oligomers (1 μ M, Bachem) for 24 h or H₂O₂ (50 μ M, Mallinckrodt Chemicals, removed after 5 min, and assessed 8 h later) unless otherwise noted. For ChIP experiments, cortical cultures (days in vitro 10–14) were used.

Dual luciferase assays were conducted on CAD cells³¹ (ATCC) that were transfected using lipofectamine (Invitrogen) with 0.5 μ g of the proximal promoter region of *Hdac2* containing the GRE consensus sequence (CAAGAAGAAAGTG GCTAC) or with the proximal promoter region without the GRE sequence subcloned into the pGL3 reporter vector (Promega) according to the manufacturer's instructions. Cells were cotransfected with 0.05 μ g of the constitutively active form of GR³², GR526 or GR_{S211A}, and treated with 1 μ M amyloid- β _{1–42} oligomers or with 50 μ M H₂O₂.

For site-directed mutagenesis, the complementary DNA for the full-length human glucocorticoid receptor (Addgene) was used and serine 211 was replaced by alanine using the QuikChange Lightning Kit (Agilent Technologies) as per the manufacturer's instructions. All constructs were verified by sequencing (Genewiz). The following mutagenesis primers were used: 5'-GTAAAGAGA CGAATGAGGCTCCTTGGAGATCAGACC-3' (forward); 5'-GGTCTGATCT CCAAGGAGCCTCATTCGTCTCTTAC-3' (reverse).

Immunohistochemistry and immunocytochemistry. Immunocytochemistry and mouse immunohistochemistry were performed as described⁷. Immunocytochemistry and immunohistochemistry on different experimental conditions were performed with the same antibody solution at the same time to assure identical staining conditions. A negative (that is, no antibody) control was included simultaneously. In brief, for immunohistochemistry, mice were perfused with 10% paraformaldehyde under deep anaesthesia (ketamine, xylazine) and their brains sectioned at 0.35 μ m thickness using a vibratome (Leica). For immunocytochemistry, cells were fixed using 4% paraformaldehyde. Slices/cells were permeabilized with 0.1% Triton X-100, blocked and incubated overnight with 0.1% Triton X-100/10% fetal bovine serum in PBS containing primary antibodies: HDAC1, HDAC2 (Abcam), HDAC3 (Santa Cruz), phospho-GR1 (S211) (Cell Signaling), GR (Abcam) or GFP (Aves Labs). Primary antibodies were visualized with Alexa-Fluor 488, Cy3 and Cy5 antibodies (Molecular Probes), neuronal nuclei with Hoechst 33342 (Invitrogen). Note that for staining of shRNA-injected animals, mCherry was visualized without staining, and only Cy2 and Cy5 secondary antibodies were used. Images were acquired using a confocal microscope (LSM 510, Zeiss) at identical settings at the highest intensity for each of the conditions. Images were quantified using ImageJ 1.42q by an experimenter blind to treatment groups, whenever possible. For each experimental condition, 20–40 representative cells per section were analysed, and the mean signal intensity was measured. To assess the specificity of the HDAC2 immunostaining, adult HDAC2^{-/-} mice were used as previously described⁷. For a specificity control of the pGR1 antibody, calf intestine phosphatase (CIP) (New England Biolabs) treatment on immunocytochemistry slides was performed as described previously³³. Calf intestine phosphatase or H₂O were applied for 40 min in Buffer 3 (New England Biolabs).

Molecular analyses. Western blots, co-immunoprecipitation, ChIP and gene expression analyses were performed as described elsewhere³⁴ with the following modifications, and expressed as fold change of the respective control conditions. For western blots, proteins were extracted using 1 \times RIPA buffer containing proteinase (complete, Roche) and phosphatase inhibitors (1 mM β -glycerophosphate, 10 mM NaF, 0.1 mM Na₃VO₄), transferred onto PVDF membranes (Biorad) and stripped using stripping buffer (Thermo Scientific). The following primary antibodies were used: acetyl-K (Cell Signaling), α - and β -tubulin, β -actin (Sigma), Cdk5 (Santa Cruz), HDAC1, HDAC2 (Abcam), HDAC3, phospho-GR1 (S211) (Cell Signaling), GR1 (Abcam), p53 (Cell Signaling) or tau (Invitrogen). Secondary antibodies were horseradish peroxidase-linked (GE Healthcare). Signal intensities were quantified using ImageJ 1.42q and normalized to values of β -actin, α - or β -tubulin. Phospho-GR1 was first normalized to GR1. Three to eight animals were used per condition. Cytoplasmic and nuclear fractionation was performed as described elsewhere³⁴.

For co-immunoprecipitation, hippocampal lysates were incubated with HDAC2 (Abcam) or IgG (Sigma) and the immunoprecipitated extracts probed for SIN3A, MTA2 (Abcam), CoRest (Millipore), LSD1 (Cell Signaling) or HDAC2. Signal intensities were quantified using ImageJ 1.42q, and normalized to input. Three or four animals were used per condition.

For ChIP, tissue samples were homogenized in cell lysis buffer containing proteinase (complete, Roche) and phosphatase inhibitors (1 mM β -glycerophosphate, 10 mM NaF, 0.1 mM Na₃VO₄) and chromatin was sonicated using a Branson Digital Sonifier with 10 rounds of 15 s at 25% power per sample on ice to 200–400 base pairs in length. For ChIP of primary cortical cultures, approximately 1 \times 10⁶ cells were crosslinked in 37% formaldehyde (Sigma), quenched with 20 \times glycine (Sigma), washed with PBS and cell lysis buffer containing both phosphatase and proteinase inhibitors (Roche) and sonicated in nuclear lysis buffer using a Vibra Cell Sonifier with 3 \times 3 pulses of 5 s at 35% power (50% duty) on ice. Sheared chromatin was immunoprecipitated with antibodies against HDAC1, HDAC2 (Abcam), HDAC3 (Santa Cruz), acetyl H2BK5, acetyl H3K14 (Abcam), acetyl H4K5, acetyl H4K12 (Millipore), phospho-GR1 (S211) (Cell Signaling) or phospho-RNA Pol II (Abcam). DNA was extracted by phenol/chloroform/isoamyl alcohol (American Bioanalytical) and subjected to quantitative PCR (Bio-Rad Thermal Cycler) using primers specific to the promoter or coding regions of the genes assayed (see Supplementary Tables 2 and 3 for primer sequences). The fluorescent signal of the amplified DNA (SYBR green, Bio-Rad) was normalized to input. Four to eight samples were used per condition.

For gene expression analysis, mRNA was extracted (Qiagen), reverse-transcribed (Invitrogen) and quantitatively amplified on a thermal cycler

(Bio-Rad) using SYBR green (Bio-Rad) and gene-specific primers (see Supplementary Table 4). The comparative C_t method³⁵ was used to examine differences in gene expression. Values were normalized to expression levels of *Gapdh*. Four to eight samples were used per condition.

Experimental manipulations. For validation of the knockdown efficacy of HDAC2 RNA interference in neurons, short hairpins targeting the open reading frame of mouse *Hdac2* mRNA from the Broad Institute's RNAi consortium shRNA library (www.broadinstitute.org/rnai/trc/lib) were packaged into lentiviral vectors³⁶ and used to infect dissociated primary mouse embryonic (gestation day 18) cortical cultures, prepared as described⁷. Primary cultures were transduced after 4 days in vitro. Cells were collected ten days after transduction, and protein levels were measured by western blotting. The target sequences for the two effective shRNAs were CCCAATGAGTTGCCATATAAT (HDAC2 shRNA 2-1, TRCN0000039395) and CGAGCATCAGACAAACGGATA (HDAC2 shRNA 2-4, TRCN0000039397).

After validation, scramble shRNA³⁷ or HDAC2-shRNA constructs were subcloned under the CaMKII-U6 promoter into the plasmid AAV entry vector fused to mCherry, tested again in mouse primary hippocampal cultures (day in vitro 7), and high titre (1×10^{12} to 4×10^{12} viral particles) adeno-associated viruses (serotype 2.5) were produced at the University of North Carolina Vector Core facility. One microlitre of shRNA-containing adeno-associated viruses was stereotactically injected into hippocampal area CA1 (anterior–posterior position -2.0 mm, medial–lateral position ± 1.6 mm, dorso–ventral -1.5 mm from Bregma) of both hemispheres at $0.1 \mu\text{l min}^{-1}$. Injection needles were left in place 5 min after injection to assure even distribution of the virus. Injections were performed 4 weeks before

behavioural testing. All infusion surgeries were performed under aseptic conditions and anaesthesia (ketamine/xylazine) in accordance with the Massachusetts Institute of Technology's Division of Comparative Medicine guidelines.

Statistics. Statistical analyses were performed using GraphPad Prism 5. One-way analyses of variance followed by Tukey post-hoc tests, or one-tailed Student's *t*-tests were used unless otherwise indicated. All data are represented as mean \pm s.e.m. Statistical significance was set at $P \leq 0.05$.

31. Qi, Y., Wang, J. K., McMillan, M. & Chikaraishi, D. M. Characterization of a CNS cell line, CAD, in which morphological differentiation is initiated by serum deprivation. *J. Neurosci.* **17**, 1217–1225 (1997).
32. Li, L. & Lindquist, S. Creating a protein-based element of inheritance. *Science* **287**, 661–664 (2000).
33. Xie, Z., Sanada, K., Samuels, B. A., Shih, H. & Tsai, L. H. Serine 732 phosphorylation of FAK by Cdk5 is important for microtubule organization, nuclear movement, and neuronal migration. *Cell* **114**, 469–482 (2003).
34. Koshibu, K. *et al.* Protein phosphatase 1 regulates the histone code for long-term memory. *J. Neurosci.* **29**, 13079–13089 (2009).
35. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_t}$. *Methods* **25**, 402–408 (2001).
36. Moffat, J. *et al.* A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124**, 1283–1298 (2006).
37. Sarbassov, D. D., Guertin, D. A., Ali, S. M. & Sabatini, D. M. Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex. *Science* **307**, 1098–1101 (2005).
38. Price, J. L. *et al.* Neuropathology of nondemented aging: presumptive evidence for preclinical Alzheimer disease. *Neurobiol. Aging* **30**, 1026–1036 (2009).

Skin infection generates non-migratory memory CD8⁺ T_{RM} cells providing global skin immunity

Xiaodong Jiang¹, Rachael A. Clark¹, Luzheng Liu¹, Amy J. Wagers², Robert C. Fuhlbrigge¹ & Thomas S. Kupper¹

Protective T-cell memory has long been thought to reside in blood and lymph nodes, but recently the concept of immune memory in peripheral tissues mediated by resident memory T (T_{RM}) cells has been proposed^{1–5}. Here we show in mice that localized vaccinia virus (VACV) skin infection generates long-lived non-recirculating CD8⁺ skin T_{RM} cells that reside within the entire skin. These skin T_{RM} cells are potent effector cells, and are superior to circulating central memory T (T_{CM}) cells at providing rapid long-term protection against cutaneous re-infection. We find that CD8⁺ T cells are rapidly recruited to skin after acute VACV infection. CD8⁺ T-cell recruitment to skin is independent of CD4⁺ T cells and interferon- γ , but requires the expression of E- and P-selectin ligands by CD8⁺ T cells. Using parabiotic mice, we further show that circulating CD8⁺ T_{CM} and CD8⁺ skin T_{RM} cells are both generated after skin infection; however, CD8⁺ T_{CM} cells recirculate between blood and lymph nodes whereas T_{RM} cells remain in the skin. Cutaneous CD8⁺ T_{RM} cells produce effector cytokines and persist for at least 6 months after infection. Mice with CD8⁺ skin T_{RM} cells rapidly cleared a subsequent re-infection with VACV whereas mice with circulating T_{CM} but no skin T_{RM} cells showed greatly impaired viral clearance, indicating that T_{RM} cells provide superior protection. Finally, we show that T_{RM} cells generated as a result of localized

VACV skin infection reside not only in the site of infection, but also populate the entire skin surface and remain present for many months. Repeated re-infections lead to progressive accumulation of highly protective T_{RM} cells in non-involved skin. These findings have important implications for our understanding of protective immune memory at epithelial interfaces with the environment, and suggest novel strategies for vaccines that protect against tissue tropic organisms.

CD8⁺ T cells have a pivotal role in antiviral immunity in target tissues^{6–9}. We infected the skin of control, CD4^{−/−}, or CD4⁺ T-cell-depleted mice with VACV and assessed VACV-specific pentamer⁺ CD8⁺ T cells¹⁰. Absence of CD4⁺ T cells did not impair either antigen-specific CD8⁺ T-cell proliferation in draining lymph nodes or subsequent accumulation in skin; in fact, the latter was enhanced (Fig. 1a, b). We then infected mice infused with OT-I (CD8⁺) and OT-II (CD4⁺) T cells with an ovalbumin-expressing VACV (VACV-Ova)¹¹. OT-I T cells are a transgenic CD8⁺ T-cell population that recognize ovalbumin residues 257–264 in the context of H-2K^b, whereas OT-II T cells are a transgenic CD4⁺ T-cell population recognizing ovalbumin residues 323–339 in the context of I-A^b (ref. 1). After skin infection, both OT-I and OT-II cells proliferated similarly in draining lymph nodes, and OT-I cells but not OT-II cells accumulated

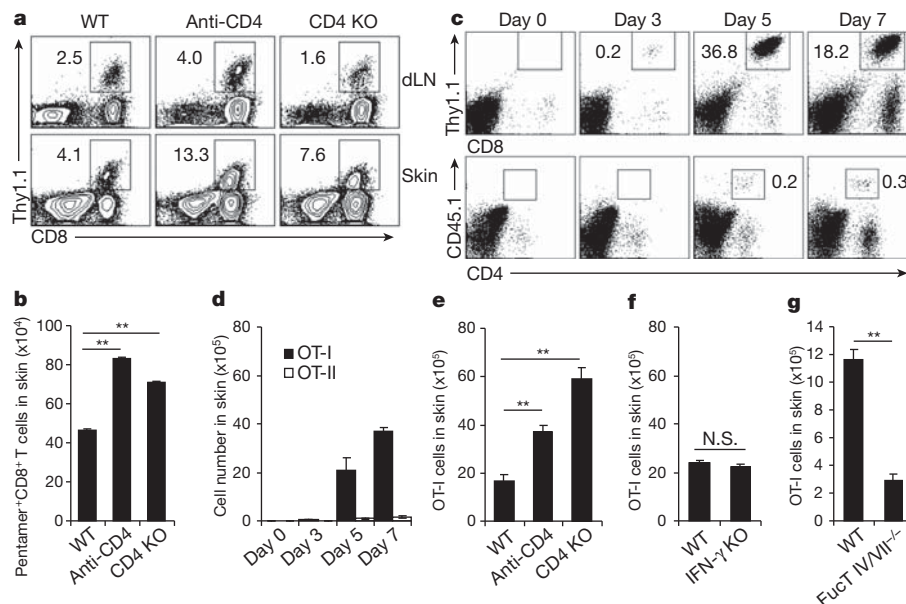


Figure 1 | CD4⁺ T cells and IFN- γ are not required for acute recruitment of CD8⁺ T cells to VACV-infected skin. **a**, **b**, Pentamer⁺ CD8⁺ T cells in draining lymph nodes (dLN) and infected skin 7 days after VACV infection. KO, knockout; WT, wild type. **a**, The percentages of pentamer⁺ CD8⁺ T cells are shown (excluding CD19⁺ cells before flow cytometry analysis). **b**, The numbers of pentamer⁺ CD8⁺ T cells in infected skin. **c**, **d**, The kinetic

infiltration of Thy1.1⁺ OT-I and CD45.1⁺ OT-II cells to infected skin after their co-transfer to naive mice. **e–g**, The numbers of OT-I cells in infected skin 7 days after infection in the absence of CD4, IFN- γ or FucT IV/VII, respectively. All data are representative of at least three independent experiments ($n = 5$ mice per time point per group). **b**, **d–g**, Error bars show standard error of the mean (s.e.m.); ** $P < 0.01$; N.S., not significant.

¹Department of Dermatology and Harvard Skin Disease Research Center, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Department of Stem Cell and Regenerative Biology, Harvard University, Howard Hughes Medical Institute, Harvard Stem Cell Institute, Joslin Diabetes Center, Boston, Massachusetts 02115, USA.

significantly in infected skin (although other $CD4^+$ T cells showed some accumulation) (Supplementary Fig. 1 and Fig. 1c, d). Interestingly, OT-I cells accumulated in infected skin efficiently in the absence of either $CD4^+$ T cells or interferon (IFN)- γ (Fig. 1e, f), in contrast to a recently reported herpes simplex virus (HSV) vaginal infection model¹². However, skin accumulation (but not lymph-node proliferation) of OT-I cells from FucT IV/VII^{-/-} mice, which cannot make E- and P-selectin ligands, was significantly impaired (Fig. 1g and Supplementary Fig. 2a). Both E- and P-selectin were significantly upregulated in VACV-infected skin (Supplementary Fig. 2b). Thus, $CD8^+$ T-cell accumulation in skin after VACV infection does not require $CD4^+$ T cells or IFN- γ , but does require expression of E- and P-selectin ligands.

Murine models of viral infections of skin and other tissues have been useful in the study of T-cell memory^{13–16}. We explored the ability of $CD8^+$ memory T cells generated by VACV infection to recirculate after resolution of the cutaneous infection. We infected the skin of mice infused with OT-I cells with VACV, and waited until complete resolution of the infection (30 days). At 30 days we could identify T_{CM} cells in lymph nodes and effector memory T (T_{EM}) cells in skin (Supplementary Fig. 3a, b). We then surgically created parabiotic pairs between the infected mice and never-infected naive mice that had not been given OT-I cells. Parabiotic pairs were maintained for 2, 4, 8, 12 and 24 weeks, at which point they were surgically separated for the analysis of VACV-specific OT-I T cells (Fig. 2a). Mice joined for 8 weeks had similar numbers of OT-I T_{CM} cells in the spleen and lymph nodes of both parabionts, indicating rapid recirculation and

equilibration of T_{CM} cells (Fig. 2b, d). However, at 2–24 weeks there were no OT-I T_{RM} cells in the skin of the unimmunized parabiont (Fig. 2c, e). These early kinetics of T_{CM} recirculation and T_{RM} non-recirculation were confirmed by parabiotic mice that received no OT-I cells, using pentamer expression to identify VACV-specific memory cells (Supplementary Fig. 4). OT-I T_{RM} cells were readily identified in the skin of previously infected parabionts and mice that had been infected in parallel but never joined. These OT-I T_{RM} cells represented a significant fraction of total skin cells (Fig. 2c), and persisted for long periods of time: 30% of skin $CD8^+$ T cells at 12 weeks, and 15–20% of skin $CD8^+$ T cells at 24 weeks. In contrast, naive mice joined to previously infected mice had no OT-I T_{RM} cells in the skin, even after 24 weeks of parabiosis (Fig. 2c, e). Thus, skin T_{RM} cells persisted in skin for at least 6 months after infection and did not recirculate appreciably. T_{RM} and T_{CM} cells were further analysed by fluorescence-activated cells sorting (FACS) for expression of CD69, CD103, E- and P-selectin ligands, and production of IFN- γ and TNF- α upon activation. A subset of T_{RM} cells expressed BCL2, and T_{RM} cells lacked CD122 and CD127 (Supplementary Fig. 5). Immunofluorescence staining showed that many T_{RM} cells localize in epidermal and follicular epithelium, as reported in skin HSV infection¹⁷, but also localize in the dermis (Fig. 2f).

To compare directly the ability of T_{CM} and T_{RM} cells to eliminate a subsequent VACV infectious challenge, we infused μ MT mice with OT-I cells and infected them through the skin with VACV-Ova. After 30 days, mice were joined parabiotically to naive μ MT mice for 8 weeks, and then surgically separated. The infected parabiont contained

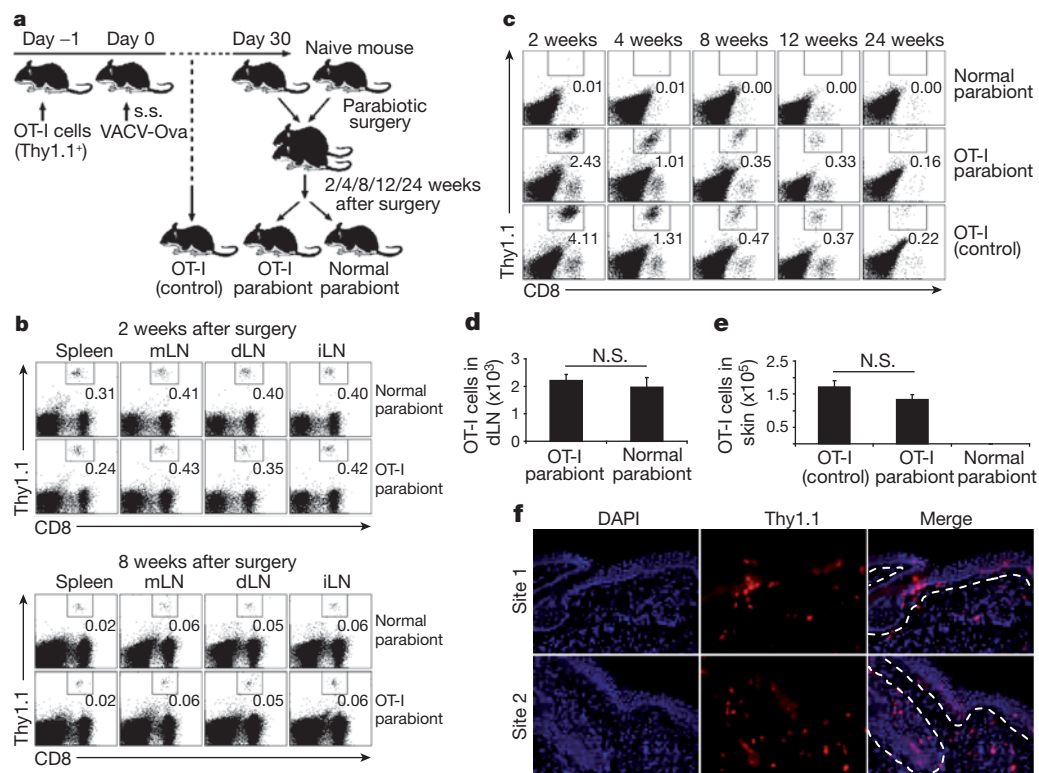


Figure 2 | $CD8^+$ T_{CM} cells recirculate quickly between parabiotic mice, but skin $CD8^+$ T_{RM} cells remain in place long term. **a**, 2×10^6 Thy1.1⁺ OT-I cells were intravenously transferred into Thy1.2⁺ mice 1 day before 2×10^6 plaque-forming units (p.f.u.) VACV-Ova skin scarification (s.s.). Thirty days after infection, OT-I-bearing mice were joined surgically with naive Thy1.2⁺ mice to create parabiotic mice. At indicated time points after surgery, parabiotic mice were separated and lymphoid tissues and skin were harvested to examine OT-I cells. **b**, The percentages of memory OT-I cells in the indicated lymphoid tissues between parabiotic mice 2 and 8 weeks after surgery were examined. mLN, mesenteric lymph node; dLN, draining lymph node; iLN, inguinal lymph

node **c**, The percentages of OT-I T_{RM} cells in the skin of parabiotic and control mice over time were also examined. **d**, The numbers of OT-I T_{CM} cells in draining lymph nodes of parabiotic mice 8 weeks after surgery. **e**, The numbers of OT-I T_{RM} cells in the skin of parabiotic and control mice 8 weeks after surgery. **b–e**, Data are representative of three independent experiments ($n = 5$ mice per time point per group). **d**, **e**, Error bars, s.e.m.; N.S., not significant. **f**, Immunofluorescence of OT-I cells in infected skin sites 45 days after infection. Sections were stained for nuclei (4',6-diamidino-2-phenylindole (DAPI), blue) and Thy1.1 (red). $n = 15$ sections from 5 mice. Original magnification, $\times 200$. Two representative sites are shown.

T_{CM} cells in both the spleen and lymph nodes and T_{RM} cells within the skin, whereas the uninfected parabiont contained T_{CM} cells only in the spleen and lymph nodes (Fig. 2). At 2 weeks after separation, the skin of these mice was challenged with VACV-Ova, and assessed 6 days later for viral load (Fig. 3a). Despite the presence of abundant circulating OT-I T_{CM} cells, uninfected parabionts cleared the virus only 30-fold more effectively than naive mice (Fig. 3b). In contrast, the infected parabiont cleared the virus completely, 10^4 -fold more effectively than

the uninfected parabiont (Fig. 3b). Viral clearance was efficient even in mice treated with FTY720, a S1P inhibitor that blocks egress of T_{CM} cells from lymph nodes into blood¹ (Fig. 3b). To show that this was not an artefact of OT-I cell transfer, we reproduced and extended the experiment in a parabiotic model not involving transfer of OT-I cells (Fig. 3c). Parabiotic pairs were separated at 4 weeks, challenged with VACV, and assessed for viral load at 6, 14 and 26 days after challenge. There were again striking differences in the ability of endogenous T_{RM}

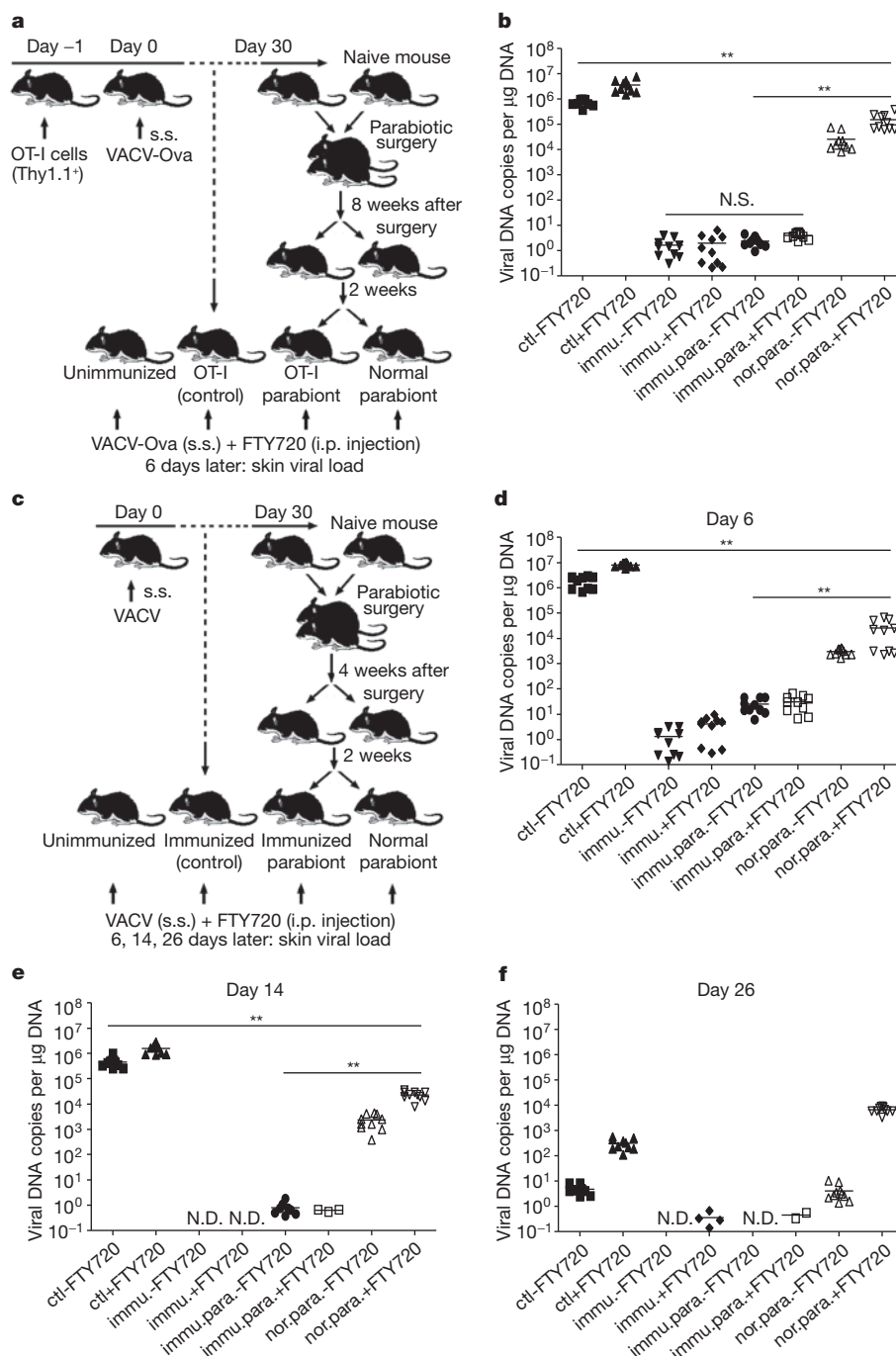


Figure 3 | Skin $CD8^+$ T_{RM} cells are superior to T_{CM} cells at protecting against re-infection. **a**, μ MT mice with OT-I transfer and VACV-Ova skin infection were used to create OT-I:normal parabiotic mice as described in Fig. 2. Eight weeks after surgery, parabiotic mice were separated. Two weeks later, separated mice were challenged with VACV-Ova on the skin. Half mice were injected daily with FTY720. **b**, Six days after challenge, skin viral load was assayed by quantitative polymerase chain reaction (qPCR). ctrl, control; immu., immunized; nor., normal; para., parabiont. **c**, In separate experiments, μ MT

mice without OT-I transfer but with VACV skin infection were used to create immunized:unimmunized parabiotic mice. Four weeks after surgery, the same VACV skin challenge protocol was applied. **d–f**, Six, fourteen or twenty-six days after challenge, viral load was assayed. The results from duplicate qPCR runs are plotted. Horizontal bars indicate the mean. Data are representative of two independent experiments ($n = 5$ mice per group). ** $P < 0.01$; N.S., not significant; N.D., not detectable.

and T_{CM} cells to mediate viral clearance at days 6 (Fig. 3d) and 14 (Fig. 3e), which began to normalize by day 26 (Fig. 3f). However, FTY720-treated T_{CM} mice (the parabiont containing T_{CM} cells but no T_{RM} cells) are unable to clear virus even 26 days after infection, a time point by which naive mice have cleared the skin infection (Fig. 3f), presumably by generating protective T_{RM} cells. Thus, although T_{CM} cells are superior to naive T cells, they are inferior to T_{RM} cells at mediating rapid viral clearance from skin.

E- and P-selectin, CCL17 and ICAM1 are expressed constitutively on the blood vessels of normal skin^{18–20} and can support entry into uninfamed skin of T_{EM} cells generated by cutaneous VACV infection. To study this phenomenon, we infected the left ears of OT-I-loaded mice with VACV-Ova and then measured the accumulation of OT-I cells in both infected (left) and uninfected (right) ears. There was measurable accumulation of OT-I cells in both ears, with similar kinetics. The absolute number of OT-I cells was always higher in the infected ear, but even 30 days after infection, OT-I cells represented a measurable fraction of all cells present in the uninfected ear (Fig. 4a). Thus, VACV skin infection generates $CD8^+ T_{RM}$ cells that distribute to distant skin sites as well as the site of infection. The accumulation of OT-I T_{RM} cells in distant skin sites is increased further after multiple sequential infections to other sites of skin (Fig. 4b–d), suggesting that skin T_{RM} cells continue to accumulate throughout skin in response to repeated cutaneous infections at distant sites.

To determine if $CD8^+ T_{RM}$ cells in distant skin sites were as protective as those at previously infected sites, we challenged μ MT mice previously infected on one ear with a second VACV skin infection on both ears either 7 or 30 days after the initial infection. FTY720 was administered to limit the contribution of T_{CM} cells (Fig. 4e). Notably, at both day 7 and day 30, T_{RM} cells in distant skin sites markedly reduced viral loads to levels comparable to those observed at the actual

site of previous infection, indicating that these distant T_{RM} cells were highly effective at rapidly eliminating virus. In contrast, viral loads in the skin of intraperitoneally (i.p.) immunized mice were between 10^3 and 10^4 higher at these time points (Fig. 4f, g). Thus, skin infection with VACV generates populations of T_{EM} cells that distribute to the entire skin surface, become T_{RM} cells, and mediate protection of the skin against re-infection with VACV in the absence of antibodies or T_{CM} cells.

T_{RM} cells have now been identified in the skin, gut, lung and brain in murine models^{1,21–23} and human subjects in both health and in the setting of skin disease^{2,4,5,24,25}. We demonstrate that after VACV viral infection through the skin, $CD8^+ T_{RM}$ cells are generated and distribute not only to the site of infection but also throughout the entire skin surface. These $CD8^+ T_{RM}$ cells produce effector cytokines, persist for many months, and are highly effective at rapidly controlling subsequent VACV skin infection. T_{RM} cells were orders of magnitude more effective than T_{CM} cells at controlling viral re-infection of the skin, at all time points examined in this study. The use of parabiotic mice allowed us to examine rigorously the tissue distribution and relative roles of T_{CM} and T_{RM} cells in VACV immune responses. The relatively minor role of $CD4^+$ T cells in VACV skin infection may reflect differences in immune responses to different viruses and/or infection of different tissues, as $CD4^+$ T cells are clearly more important in HSV infection^{12,26}. Moreover, $CD4^+$ T_{RM} cells predominate in human skin²⁴ and lung²⁵, and are enriched for memory cells specific for pathogens encountered through those tissues^{25,27,28}.

Pathogens typically invade the host through epithelial interfaces with the environment. Our studies suggest that T_{EM} cells generated as a result of epithelial tissue infections accumulate as T_{RM} cells at both sites of infection as well as at distant sites within the same epithelial tissue, providing broad and long-lived protective T-cell immunity against

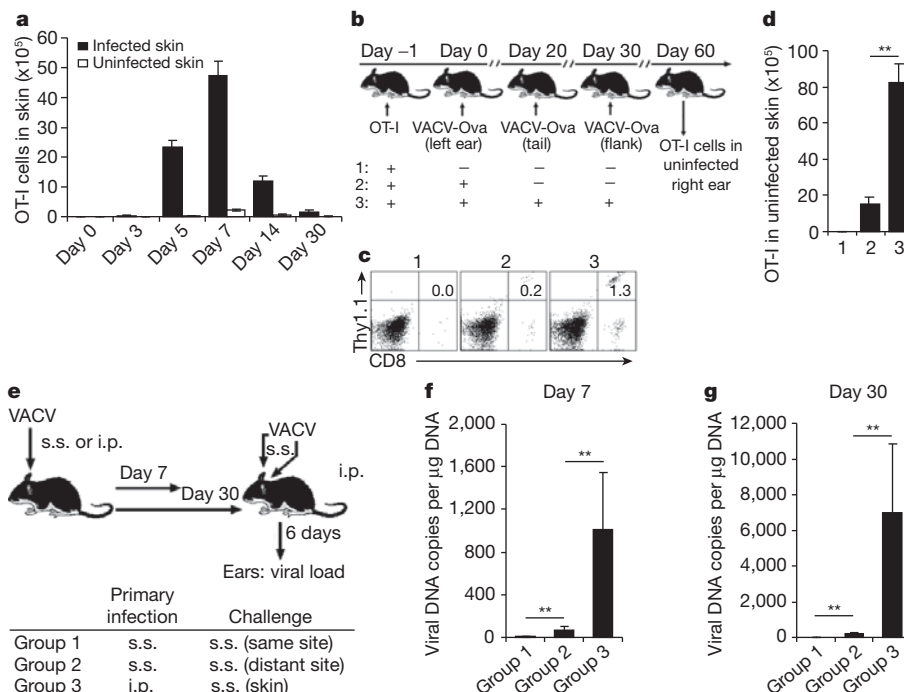


Figure 4 | Skin $CD8^+ T_{RM}$ cells also accumulate in unimmunized site after skin infection and are highly effective at eliminating virus. **a**, 2×10^6 Thy1.1⁺ OT-I cells were intravenously transferred into Thy1.2⁺ mice 1 day before 2×10^6 p.f.u. VACV-Ova skin scarification. The numbers of OT-I cells in infected and uninfected skin were enumerated over time. **b**, The left ears of OT-I-infused B6 mice were infected with VACV-Ova. Twenty and thirty days later, some of the mice were challenged with 2×10^6 p.f.u. VACV-Ova on the tail and flank skin, respectively. All infection routes were by skin scarification. Thirty days after challenge, Thy1.1⁺ OT-I cells in uninfected right ears were

examined. **c**, The percentages of Thy1.1⁺ OT-I cells in uninfected right ears. **d**, The numbers of Thy1.1⁺ OT-I cells in uninfected right ears. Error bars, s.e.m. **e**, μ MT mice were infected with VACV either by skin scarification on one ear or i.p. injection. Seven or thirty days later, mice were challenged with VACV on both ears via skin scarification (s.s.). Daily i.p. injection of FTY720 was performed. **f**, **g**, Six days later, viral loads in both infected and uninfected ears were measured. The mean and s.d. from triplicate qPCR runs are plotted. Data are representative of three independent experiments ($n = 5$ mice per group). ** $P < 0.01$.

re-infection. A more complete understanding of T_{RM}-cell-mediated immune memory should enhance our understanding of adaptive immunological memory, influence rational vaccine design, and illuminate the pathophysiology of human T-cell-mediated diseases.

METHODS SUMMARY

VACV skin infection. 2×10^6 p.f.u. of recombinant VACV Western Reserve strain or VACV expressing full-length ovalbumin (VACV-Ova) was used for epicutaneous infection by skin scarification or i.p. injection, as described previously²⁹.

Parabiotic mice. Well-matched (sex and age) mouse partners were anaesthetized to full muscle relaxation with ketamine and xylazine ($10 \mu\text{g g}^{-1}$) or 2.5% avertin ($15 \mu\text{l g}^{-1}$) by i.p. injection. The corresponding lateral aspects of mice were shaved and the excess hair was wiped off with alcohol prep pad. The disinfection was performed by wiping with betadine solution and 70% ethanol three times. Matching skin incisions were made from the olecranon to the knee joint of each mouse, and the subcutaneous fascia was bluntly dissected to create about 0.5 cm of free skin. The olecranon and knee joints were attached by a single 5-0 silk suture and tie, and the dorsal and ventral skins were approximated by staples or continuous suture. Betadine solution was used to cover the full length of the dorsal and ventral incision. The mice were then kept on heating pads and continuously monitored until recovery. For analgesic treatment, mice were injected subcutaneously with $2.5 \mu\text{g g}^{-1}$ flunixin every 8–12 h for 48 h after the operation.

Cell isolation from skin. Before harvest, skin hair was removed using Nair Hair Remover. Skin samples were then chopped into small fragments and incubated in Hanks balanced salt solution (HBSS) supplemented with 1 mg ml^{-1} collagenase A (Roche) and $40 \mu\text{g ml}^{-1}$ DNase I (Roche) at 37°C for 30 min. After filtrating through a $70\text{-}\mu\text{m}$ nylon cell strainer, cells were collected and washed thoroughly with cold PBS before staining.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 June 2011; accepted 12 January 2012.

Published online 29 February 2012.

- Liu, L. *et al.* Epidermal injury and infection during poxvirus immunization is crucial for the generation of highly protective T cell-mediated immunity. *Nature Med.* **16**, 224–227 (2010).
- Clark, R. A. *et al.* Skin effector memory T cells do not recirculate and provide immune protection in alemtuzumab-treated CTCL patients. *Sci. Transl. Med.* **4**, 117 (2012).
- Sheridan, B. S. & Lefrançois, L. Regional and mucosal memory T cells. *Nature Immunol.* **13**, 485–491 (2011).
- Boyman, O. *et al.* Spontaneous development of psoriasis in a new animal model shows an essential role for resident T cells and tumor necrosis factor- α . *J. Exp. Med.* **199**, 731–736 (2004).
- Conrad, C. *et al.* $\alpha_1\beta_1$ integrin is crucial for accumulation of epidermal T cells and the development of psoriasis. *Nature Med.* **13**, 836–842 (2007).
- Bevan, M. J. Helping the CD8⁺ T-cell response. *Nature Rev. Immunol.* **4**, 595–602 (2004).
- Antia, R., Ganusov, V. V. & Ahmed, R. The role of models in understanding CD8⁺ T-cell memory. *Nature Rev. Immunol.* **5**, 101–111 (2005).
- Kaech, S. M. & Wherry, E. J. Heterogeneity and cell-fate decisions in effector and memory CD8⁺ T cell differentiation during viral infection. *Immunity* **27**, 393–405 (2007).
- Lefrançois, L. & Obar, J. J. Once a killer, always a killer: from cytotoxic T cell to memory cell. *Immunol. Rev.* **235**, 206–218 (2010).
- Freyschmidt, E. J. *et al.* Skin inflammation arising from cutaneous regulatory T cell deficiency leads to impaired viral immune responses. *J. Immunol.* **185**, 1295–1302 (2010).
- Sanz, P. & Moss, B. Identification of a transcription factor, encoded by two vaccinia virus early genes, that regulates the intermediate stage of viral gene expression. *Proc. Natl Acad. Sci. USA* **96**, 2692–2697 (1999).
- Nakanishi, Y., Lu, B., Gerard, C. & Iwasaki, A. CD8⁺ T lymphocyte mobilization to virus-infected tissue requires CD4⁺ T-cell help. *Nature* **462**, 510–513 (2009).
- Wakim, L. M., Waithman, J., van Rooijen, N., Heath, W. R. & Carbone, F. R. Dendritic cell-induced memory T cell activation in nonlymphoid tissues. *Science* **319**, 198–202 (2008).
- Masopust, D., Vezys, V., Marzo, A. L. & Lefrançois, L. Preferential localization of effector memory cells in nonlymphoid tissue. *Science* **291**, 2413–2417 (2001).
- Lefrançois, L. Development, trafficking, and function of memory T-cell subsets. *Immunol. Rev.* **211**, 93–103 (2006).
- Klonowski, K. D. *et al.* Dynamics of blood-borne CD8 memory T cell migration *in vivo*. *Immunity* **20**, 551–562 (2004).
- Gebhardt, T. *et al.* Different patterns of peripheral migration by memory CD4⁺ and CD8⁺ T cells. *Nature* **477**, 216–219 (2011).
- Chong, B. F., Murphy, J. E., Kupper, T. S. & Fuhlbrigge, R. C. E-selectin, thymus- and activation-regulated chemokine/CCL17, and intercellular adhesion molecule-1 are constitutively coexpressed in dermal microvessels: a foundation for a cutaneous immunosurveillance system. *J. Immunol.* **172**, 1575–1581 (2004).
- Weninger, W. *et al.* Specialized contributions by $\alpha(1,3)$ -fucosyltransferase-IV and FucT-VII during leukocyte rolling in dermal microvessels. *Immunity* **12**, 665–676 (2000).
- Jiang, X., Campbell, J. J. & Kupper, T. S. Embryonic trafficking of $\gamma\delta$ T cells to skin is dependent on E/P selectin ligands and CCR4. *Proc. Natl Acad. Sci. USA* **107**, 7443–7448 (2010).
- Gebhardt, T. *et al.* Memory T cells in nonlymphoid tissue that provide enhanced local immunity during infection with herpes simplex virus. *Nature Immunol.* **10**, 524–530 (2009).
- Masopust, D. *et al.* Dynamic T cell migration program provides resident memory within intestinal epithelium. *J. Exp. Med.* **207**, 553–564 (2010).
- Wakim, L. M., Woodward-Davis, A. & Bevan, M. J. Memory T cells persisting within the brain after local infection show functional adaptations to their tissue of residence. *Proc. Natl Acad. Sci. USA* **107**, 17872–17879 (2010).
- Clark, R. A. *et al.* The vast majority of CLA⁺ T cells are resident in normal skin. *J. Immunol.* **176**, 4431–4439 (2006).
- Purwar, R. *et al.* Resident memory T cells (T_{RM}) are abundant in human lung: diversity, function, and antigen specificity. *PLoS ONE* **6**, e16245 (2011).
- Lund, J. M., Hsing, L., Pham, T. T. & Rudensky, A. Y. Coordination of early protective immunity to viral infection by regulatory T cells. *Science* **320**, 1220–1224 (2008).
- Román, E. *et al.* CD4 effector T cell subsets in the response to influenza: heterogeneity, migration, and function. *J. Exp. Med.* **196**, 957–968 (2002).
- Romani, L. Immunity to fungal infections. *Nature Rev. Immunol.* **11**, 275–288 (2011).
- Liu, L., Fuhlbrigge, R. C., Karibian, K., Tian, T. & Kupper, T. S. Dynamic programming of CD8⁺ T cell trafficking after live viral immunization. *Immunity* **25**, 511–520 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Tian, R. Purwar and Q. Zhan for technical assistance. We thank J. J. Campbell for discussion of the project. This work was supported by National Institutes of Health (NIH) grants R01AI041707, R37AI025082 and TR01 AI097128 to T.S.K.

Author Contributions X.J. and T.S.K. designed research; X.J. performed research; L.L. helped to establish the VACV skin scarification model; A.J.W. helped to create parabiotic mice; X.J., R.A.C., R.C.F., L.L. and T.S.K. analysed data; and X.J., R.A.C. and T.S.K. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.S.K. (tkupper@partners.org).

METHODS

Mice. C57BL/6, CD4^{-/-}, IFN- γ ^{-/-} and μ MT mice were purchased from The Jackson Laboratory. Thy1.1⁺ Rag^{-/-} OT-I, CD45.1⁺ OT-II, FucT IV/VII^{-/-} mice were housed at the animal facility of Harvard Institute of Medicine, Harvard Medical School. Thy1.1⁺ Rag^{-/-} OT-I mice were crossed with FucT IV/VII^{-/-} mice to yield Thy1.1⁺ FucT IV/VII^{-/-} OT-I mice. OT-I and OT-II are T-cell-receptor-transgenic mice recognizing chicken ovalbumin residues 257–264 in the context of H-2K^b and 323–339 in the context of I-A^b, respectively. Animal experiments were performed in accordance with the guidelines set out by the Center for Animal Resources and Comparative Medicine at Harvard Medical School.

Viruses and infections. Recombinant VACV (Western Reserve strain) and VACV-Ova were originally obtained from B. Moss (NIH) and grown on HeLa cells. Viral titres were determined using CV-1 cells and 2×10^6 p.f.u. of VACV or VACV-Ova was used for epicutaneous infection by skin scarification or i.p. injection, as described previously²⁹.

Parabiotic mice. Parabiosis surgery was performed as described elsewhere³⁰ with some modifications. Briefly, sex- and age-matched mouse partners were anaesthetized to full muscle relaxation with ketamine and xylazine ($10 \mu\text{g g}^{-1}$) or with 2.5% avertin ($15 \mu\text{l g}^{-1}$) by i.p. injection. The corresponding lateral aspects of mice were shaved and the excess hair was wiped off with alcohol prep pad. After skin disinfection by wiping with betadine solution and 70% ethanol three times, two matching skin incisions were made from the olecranon to the knee joint of each mouse, and the subcutaneous fascia was bluntly dissected to create about 0.5 cm of free skin. The olecranon and knee joints were attached by a single 5-0 silk suture and tie, and the dorsal and ventral skins were approximated by staples or continuous suture. Betadine solution was used to cover the full length of the dorsal and ventral incision. The mice were then kept on heating pads and continuously monitored until recovery. $2.5 \mu\text{g g}^{-1}$ flunixin was used for analgesic treatment by subcutaneous injection every 8–12 h for 48 h after the operation. After an interval of the indicated weeks, parabiotic mice were surgically separated by a reversal of the above procedure for the next experiments.

Adoptive transfer and T-cell depletion. Lymph nodes were collected from the naive female Thy1.1⁺ Rag^{-/-} OT-I, CD45.1⁺ OT-II, or Thy1.1⁺ FucT IV/VII^{-/-} OT-I mice at the age of 6–8 weeks. OT-I or OT-II cells were purified by magnetic cell sorting using mouse CD8 α ⁺ or CD4⁺ T-cell isolation kit (Miltenyi Biotec), respectively. 2×10^6 isolated OT-I and/or OT-II cells were then intravenously transferred to female recipient mice. In some experiments, OT-I and OT-II cells were labelled with carboxyfluorescein succinimidyl ester (CFSE) before co-transfer. To deplete CD4⁺ T cells *in vivo*, the recipient mice were injected i.p. with $500 \mu\text{g}$ anti-CD4 (GK1.5) in $100 \mu\text{l}$ PBS 4 days and 1 day before and on day 2 and 5 after infection.

Preparation of cell suspensions. Lymph nodes and spleen were harvested and mashed through a $70\text{-}\mu\text{m}$ nylon cell strainer to prepare cell suspensions. Red blood cells were lysed using lysing buffer. Skin tissue was removed after hair removal, chopped into small fragments and incubated in Hanks balanced salt solution (HBSS) supplemented with 1 mg ml^{-1} collagenase A and $40 \mu\text{g ml}^{-1}$ DNase I at 37°C for 30 min. After filtrating through a $70\text{-}\mu\text{m}$ nylon cell strainer, cells were collected and washed thoroughly with cold PBS before staining.

Intracellular cytokine detection. The infected skin from memory OT-I-bearing mice was harvested at 35 days after infection, and single-cell suspensions were prepared as described above. In some cases, 2 months after skin scarification with VACV-Ova, mice transferred with 2×10^6 OT-I cells were i.p. challenged with VACV-Ova. Five days after challenge, splenocytes were prepared. Red blood cells

were lysed using lysing buffer. Cells were then incubated with $2 \mu\text{g ml}^{-1}$ SINFEKL peptide of ovalbumin at the presence of Brefeldin A for 7 h. Fc receptors were blocked with CD16/CD32 monoclonal antibodies and intracellular IFN- γ , TNF- α and IL-2 staining was performed using Intracellular Cytokine Detection Kits (BD Bioscience) before flow cytometry.

Determination of viral load. Mice were challenged with 2×10^6 p.f.u. VACV or VACV-Ova on the skin. In some cases, mice were simultaneously injected (i.p.) with $1 \mu\text{g g}^{-1}$ FTY720 each day. At indicated time points, viral load of skin was examined by quantitative real-time PCR, as described previously²⁹.

Isolation of mRNA and real-time PCR. Total RNA was extracted from homogenized skin tissue and cDNA was generated with iScript cDNA synthesis kit (Bio-Rad). Bio-Rad iCycler iQ Real-Time PCR Detection System (Bio-Rad) was used with the following settings: 45 cycles of 15 s of denaturation at 95°C , and 1 min of primer annealing and elongation at 60°C . Real-time PCR was done with $1 \mu\text{l}$ cDNA plus $12.5 \mu\text{l}$ of $2 \times$ iQ SYBR Green Supermix (Bio-Rad) and $0.5 \mu\text{l}$ ($10 \mu\text{M}$) specific primers: mouse E-selectin 1 (5'-GGACACCACAAAATCCAGTCTG-3') and mouse E-selectin 2 (5'-TCGACAGGAGAAGTCACTG-3'); mouse P-selectin 1 (5'-AAGATGCCTGGCTACTGGACAC-3') and mouse P-selectin 2 (5'-CAAGAGGCTGAACGCAGGTCAT-3'); mouse β -actin 1 (5'-CA TTGCTGACAGGATGCAGAAGG-3') and mouse β -actin 2 (5'-TGCTGGAA GGTGGACAGTGAGG-3'). All samples were run in duplicate and fold change of gene expression was calculated using the reference sample (naive skin).

Antibodies and flow cytometry. The following anti-mouse antibodies were obtained from BD PharMingen: CD8a (53-6.7), Thy1.1 (OX-7), CD4 (L3T4), CD45.1 (A20), CD19 (1D3), CD16/CD32 (2.4G2), CD44 (IM7), CD62L (MEL-14), CD69 (H1.2F3), CD103 (M290), CD122 (TM-Beta 1), BCL2 (A19-3), α 4 β 7 (DATK32), IFN- γ (XMG1.2), TNF- α (MP6-XT22), IL-2 (JES6-5H4). Fluorescence-conjugated anti-mouse CD127 (A7R34) and PD-1 (RMP1-30) were purchased from eBioscience. PE-conjugated B8R₂₀₋₂₇/H-2K^b pentamers were obtained from ProImmune Ltd and pentamer⁺ CD8⁺ T-cell staining was performed according to the protocol provided by the company. E- or P-selectin ligand expression was examined by incubating cells with rmE-Selectin/Fc Chimera or rmP-Selectin/Fc Chimera (R&D System) in conjunction with APC-conjugated F(ab')₂ fragments of goat anti-human IgG F(c) antibody (Jackson Immuno-research). Dead cells were excluded using 7-AAD staining. Data were analysed on FACSCanto Flow Cytometer using FACSDiva software.

Immunofluorescence microscopy. One centimetre of tail containing the infected skin site from OT-I-bearing mice was harvested and deboned at 45 days after skin scarification with VACV-Ova. The skin was collected and embedded in OCT (Tissue-Tek; Sakura) and frozen at -80°C . Skin sections were performed on a cryostat (Leica CM1850 UV) at $6\text{-}\mu\text{m}$ thickness and air-dried for 6–8 h. Sections were then fixed in -20°C acetone for 5 min, rehydrated with PBS, and blocked with 2% FCS in PBS for 15 min at room temperature (20°C). After staining with Thy1.1-PE (1:200) in PBS for 1 h at room temperature, sections were rinsed three times (for 5 min each time) with TBS-tween 20 by shaking and mounted with DAPI Fluoromount-G (Southern Biotech). Images were acquired with a Nikon Eclipse E600 microscope and processed using SPOTSOFTWARE (version 4.6, Diagnostic Instruments).

Statistical analysis. Statistical significance in values between experimental groups was determined by one-way analysis of variance (ANOVA) followed by Tukey post-test. $P < 0.05$ was considered statistically significant.

- Wagers, A. J., Sherwood, R. I., Christensen, J. L. & Weissman, I. L. Little evidence for developmental plasticity of adult hematopoietic stem cells. *Science* **297**, 2256–2259 (2002).

Preservation of organic matter in sediments promoted by iron

Karine Lalonde¹, Alfonso Mucci², Alexandre Ouellet¹ & Yves G  linas¹

The biogeochemical cycles of iron and organic carbon are strongly interlinked. In oceanic waters, organic ligands have been shown to control the concentration of dissolved iron¹. In soils, solid iron phases shelter and preserve organic carbon², but the role of iron in the preservation of organic matter in sediments has not been clearly established. Here we use an iron reduction method previously applied to soils³ to determine the amount of organic carbon associated with reactive iron phases in sediments of various mineralogies collected from a wide range of depositional environments. Our findings suggest that 21.5 ± 8.6 per cent of the organic carbon in sediments is directly bound to reactive iron phases. We further estimate that a global mass of $(19\text{--}45) \times 10^{15}$ grams of organic carbon is preserved in surface marine sediments as a result of its association with iron⁴. We propose that these associations between organic carbon and iron, which are formed primarily through co-precipitation and/or direct chelation, promote the preservation of organic carbon in sediments. Because reactive iron phases are metastable over geological timescales, we suggest that they serve as an efficient 'rusty sink' for organic carbon, acting as a key factor in the long-term storage of organic carbon and thus contributing to the global cycles of carbon, oxygen and sulphur⁵.

Evidence of interactions between iron and organic carbon in marine sediments was reported nearly 40 yr ago, where concentrations of iron and organic carbon were found to co-vary⁶. Because both iron and organic carbon are commonly associated with clay mineral surfaces, it was simply stated that "where there is more deposited fine-grained material with high surface area for adsorption, we find more organic matter and more Fe"⁶. It is still not clear whether this correlation stems from the strong affinity of both species for solid surfaces or whether it reflects enhanced preservation of organic carbon by iron. Iron's preservative effect on organic matter was previously demonstrated in laboratory studies^{7,8}, which reported that the presence of iron-rich solid substrates or the formation of organoferric complexes hampers microbial degradation of simple organic compounds. Iron also imparts a protective effect to organic carbon in soil systems², but this preservation mechanism has never been explored in sediments.

In modern sediments, reactive iron phases (defined here as the solid iron phases that are reductively dissolved by sodium dithionite) are typically found as nanospheres of goethite <10 nm in diameter^{9,10}. These phases accumulate or are formed within the oxic sediment layer through oxidation and precipitation of dissolved iron(II) produced during weathering and diagenetic recycling within the sediment¹¹. Over time, reactive iron phases become more crystalline, resulting in reduced surface area, reactivity and solubility. Crystallization is, however, hindered by the active diagenetic recycling of iron¹² and by coating of iron phases by organic matter¹³. Accordingly, reactive iron phases have been shown to survive in sediments for hundreds of thousands of years¹⁴.

We examined sediments collected from a wide range of environments, including fresh waters, estuaries, river deltas, continental margins

and the deep sea and encompassing various depositional environments and mineralogies. These samples include organic-carbon-rich, sulphidic Black Sea sediments and organic-carbon-rich sediments from dioxygen-deficient zones along the Indian and Mexican (sampling station 306) margins. Also included are sediments from the Arabian Sea, the Saanish Inlet and a boreal lake (Lake Brock), which have a productivity-driven seasonal pattern of dioxygen-deficient waters. Estuarine, deltaic and margin deposits accumulating below well-oxygenated waters of the Arctic margin, the St Lawrence estuary and gulf, the Mexican margin (stations 303–305), the Eel River basin, the Washington coast and the adjacent Columbia River delta are also examined along with pelagic sediments from the Southern Ocean, the Santa Barbara basin (station M) and the equatorial Pacific Ocean. This sample set comprises fresh water, estuarine and marine clastic sediments, carbonate and siliceous oozes, as well as pelagic red clay sediments.

We focused on determining the amount of organic carbon associated with reactive iron phases by applying the citrate–dithionite iron reduction method of ref. 15, which simultaneously dissolves from the sediment matrix all solid reactive iron phases and the organic carbon associated with these phases (OC-Fe). The reduction reaction is conducted at circumneutral pH using sodium bicarbonate as a buffer, thus preventing the hydrolysis of organic matter as well as its protonation and re-adsorption onto sediment particles, which occur under acidic conditions. Whereas the extraction of the same samples with artificial sea water released a negligible fraction of the total organic carbon (less than 3%; results not shown), samples treated under the same experimental conditions after substituting trisodium citrate (complexing agent) and sodium dithionite (reducing agent) for sodium chloride (equivalent ionic strength) released on average $7.2 \pm 5.4\%$ of the total organic carbon (Supplementary Table 2). Because the organic carbon released in these control experiments is not associated with iron, results of individual control experiments were subtracted from the amount of organic carbon released from the dithionite extractions (see Supplementary method for results and a discussion on contamination and specificity for the OC-Fe fraction).

We determined that for all sediments tested, an average of $20.5 \pm 7.8\%$ of the total organic carbon is directly associated with iron, with the highest OC-Fe concentrations in the uppermost sediment layers, where most of the reactive iron phases accumulate (Fig. 1). Considering organic carbon burial within different depositional settings—deltaic and continental margin sediments respectively account for 44% and 45% of global organic carbon burial, whereas pelagic sediments and high productivity zones, including anoxic basins, respectively account for 5% and 6% (ref. 16)—we estimate that the global pool of organic carbon specifically associated with iron corresponds to $21.5 \pm 8.6\%$ of the total sedimentary organic carbon, or $(19\text{--}45) \times 10^{15}$ g of organic carbon. Even in mature sediments (1,000–1,500 yr old), 23–27% of the total organic carbon remains bound to reactive iron oxide phases, suggesting that the strong association between iron and organic carbon may inhibit microbial organic carbon degradation and enhance organic carbon preservation.

¹GEOTOP and Department of Chemistry and Biochemistry, Concordia University, 7141 Sherbrooke Street West, Montreal, Quebec H4B 1R6, Canada. ²GEOTOP and Earth and Planetary Sciences, McGill University, 3450 University Street, Montreal, Quebec H3A 2A7, Canada.

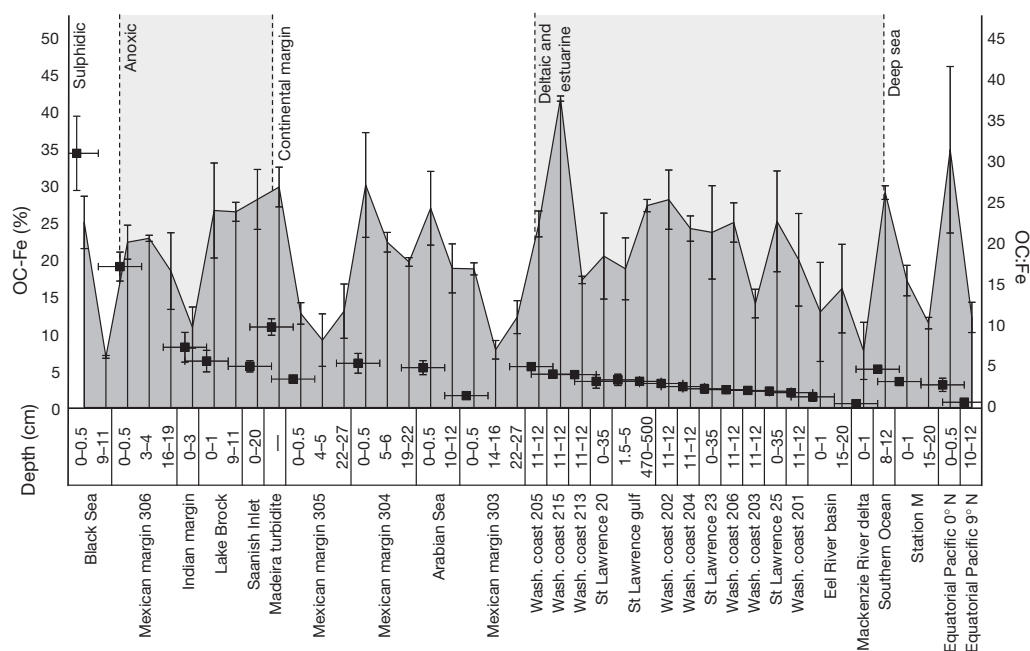


Figure 1 | Control-corrected percentage of the total sediment organic carbon bound to reactive iron phases. The OC:Fe considered here (black line) is that dislodged from the sediment during the reductive dissolution of reactive iron oxides. Depth intervals are indicated below the x axis, along with the corresponding geographical area and, if relevant, sampling station. Molar OC:Fe ratios of the uppermost surface sediment layer are also shown (black squares). The iron reduction was carried out following the method of ref. 15 without adding agents that promote flocculation of the dissolved organic matter after the reduction step. Error bars, s.d.; $n = 12-15$ for the St Lawrence samples and $n = 3$ for the others.

In agreement with the calculations in ref. 3, our measurements reveal that reactive iron phases do not provide sufficient surface area ($<5\%$ of the total surface area of sediments; Supplementary Table 3) for adsorption of the entire OC-Fe pool onto iron oxides. As an alternative explanation, we propose the existence of largely organic OC-Fe macromolecular structures that are dissolved and dislodged from the sediment during iron reduction. Transmission electron microscopy studies describe sedimentary organic matter as “discrete, discontinuous blebs” that adhere to the surface of sediment clay particles¹⁷. These blebs are consistent with our proposed structure of OC-Fe, as are the findings of ref. 16, where it was reported that sedimentary organic matter is not spread evenly over clay particles but covers only about 15% of the surface. We believe that iron or iron oxides are critical in providing cohesion to these macromolecular structures, possibly fixing them to clay particles through strong covalent bonds.

Calculations indicate that simple sorption of organic matter on reactive iron oxide surfaces results in a maximum molar ratio of

organic carbon to iron (OC:Fe) of 1.0 for the co-extracted species, based on the maximal sorption capacity of reactive iron oxides for natural organic matter³. However, co-precipitation and/or chelation of organic compounds with iron generates low-density, organic-rich structures with OC:Fe ratios between 6 and 10 (ref. 3). According to the results of our dithionite extractions, typical continental margin sediments overlain by oxic bottom waters yield an average OC:Fe ratio of 4.0 ± 2.8 (Supplementary Table 3), which greatly exceeds the maximum sorption capacity of iron oxides but is consistent with the formation of OC-Fe chelates. These chelates are predominantly organic structures that probably resemble those depicted by the ‘onion model’¹⁸, where organic molecules are ‘glued’ together by iron ions or nanophases of iron oxides. The formation of such chelates from solution is possible when the molar porewater OC:Fe ratio is approximately 10 (refs 19, 20). This molar ratio is typically observed in anoxic sediment pore waters such as that in the St Lawrence estuary (K.L., unpublished data) and in the nearby Saguenay fjord²¹. The diffusion of dissolved iron(II) from

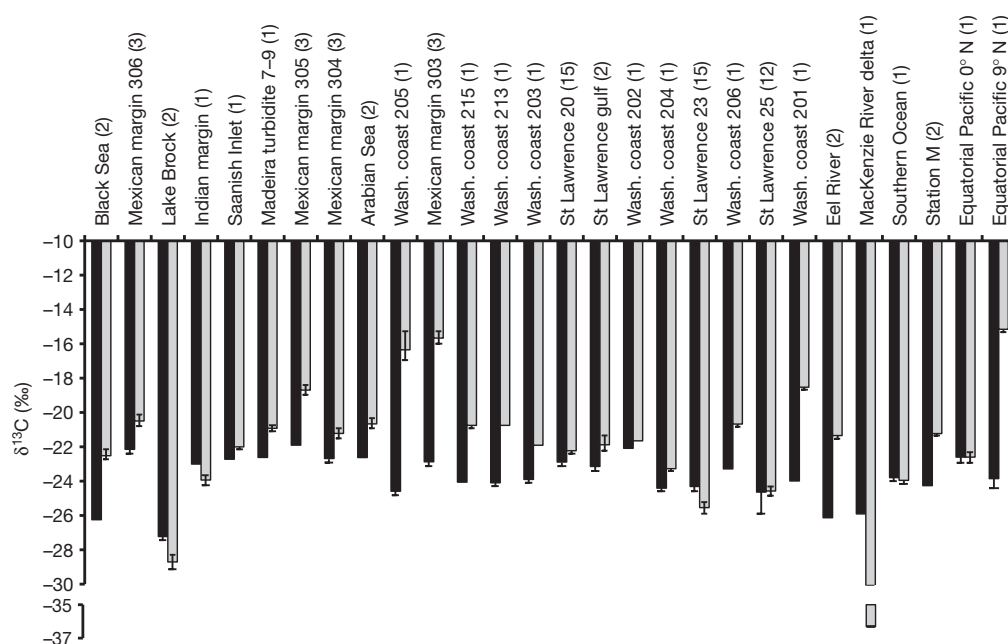


Figure 2 | Carbon isotopic signatures of non-iron-bound and iron-bound organic carbon for all sediment samples. The samples (non-iron-bound, black; iron-bound, grey) were depth-integrated whenever possible; the number of depth intervals integrated is indicated in parentheses above the sample name. $\delta^{13}\text{C} = (^{13}\text{C}/^{12}\text{C})_{\text{sample}} / (^{13}\text{C}/^{12}\text{C})_{\text{VPDB}} - 1$; VPDB, Vienna Pee Dee Belemnite. Error bars, s.d.; $n = 12-15$ for the St Lawrence samples and $n = 3$ for the others.

anoxic to surficial oxic sediments would trigger the oxidation of iron(II) to iron(III) and the formation of very stable organic complexes^{22,23} (the formation constant ranges from 10^{14} M^{-1} for natural dissolved organic carbon to 10^{52} M^{-1} for siderophores).

Sediments bathed by oxygen-depleted bottom waters, such as in the Black Sea, the Mexican margin (station 306) and the Indian margin, host structures with high OC:Fe ratios (7 to 32). These organometallic structures seem to be particularly stable under anaerobic conditions and survive degradation. By contrast, in oxic environments, the organic lining of these structures is progressively degraded, reducing the OC:Fe ratio to values observed in typical continental margin sediments (Fig. 1). Long exposure to oxic conditions increases the fraction of the total sedimentary organic carbon pool that is tightly adsorbed to particle surfaces²⁴, owing to the preferential degradation of organic structures that are more loosely attached to the clay mineral matrix, such as the OC-Fe chelates. Very long exposure to oxic conditions results in the very low OC:Fe ratio observed at the deep-sea equatorial Pacific site (0.36; Fig. 1).

We also analysed the isotopic composition ($\delta^{13}\text{C}$ and $\delta^{15}\text{N}$) and elemental composition (molar ratio of carbon to nitrogen) of the bulk organic matter and the iron-associated organic carbon fractions of all sediment samples. In most cases, we find that OC-Fe is enriched in ^{13}C ($\delta^{13}\text{C}$ increases by $1.7 \pm 2.8\%$; Fig. 2) and nitrogen (C:N decreases by 1.7 ± 2.8) relative to the rest of the sedimentary organic carbon pool, whereas $\delta^{15}\text{N}$ shows little or no fractionation (Supplementary Figs 1 and 2). Natural organic compounds rich in ^{13}C include proteins and carbohydrates²⁵, which are rich in nitrogen and/or oxygen functionalities that favour the formation of inner-sphere complexes with iron. The preferential binding of such highly labile organic compounds to iron may explain why reactive organic compounds can be preserved in sediments whereas other, more refractory, molecules are degraded⁴.

Our findings have far-reaching implications for our understanding of organic matter cycling in sediments. First, the protection mechanism described above, which preferentially shields ^{13}C - and nitrogen-rich organic compounds from microbial degradation, could help explain a phenomenon that has puzzled organic geochemists for decades: the replacement, seaward of river mouths, of terrigenous organic matter from sediments by compounds bearing a more marine isotopic and elemental signature²⁶. Our data also show that the traditional sorptive stabilization mechanism, which proposes that clay particles have a preservative effect on organic matter through direct adsorption on their surfaces^{4,27,28}, does not describe accurately the mode of stabilization for all organic compounds in sediments. Although more work is needed to elucidate the exact nature of interactions between organic carbon and iron, our data suggest that direct chelation or co-precipitation of macromolecular OC-Fe structures has a significant role. Finally, our results reveal that $21.5 \pm 8.6\%$ of the organic carbon buried in surface marine sediments ($150 \times 10^{15} \text{ g}$ of organic carbon⁴), or a global mass of $(19\text{--}45) \times 10^{15} \text{ g}$ of organic carbon, is preserved as a result of its intimate association with reactive iron phases. Assuming that our estimate also applies to organic carbon locked in the sedimentary rock reservoir ($150,000 \times 10^{18} \text{ g}$ of organic carbon⁴), iron-associated organic carbon would account for $(1,900\text{--}4,500) \times 10^{18} \text{ g}$ of organic carbon, or roughly 2,900–6,800 times the amount in the atmospheric carbon pool. Hence, reactive iron phases serve as an extremely efficient ‘rusty sink’ for organic carbon and are a key factor in the long-term storage of organic carbon and the global cycles of carbon, oxygen and sulphur.

Received 22 February 2011; accepted 10 January 2012.

1. Johnson, K. S., Gordon, R. M. & Coale, K. H. What controls dissolved iron in the world ocean? *Mar. Chem.* **57**, 137–161 (1997).
2. Kaiser, K. & Guggenberger, G. The role of DOM sorption to mineral surfaces in the preservation of organic matter in soils. *Org. Geochem.* **31**, 711–725 (2000).

3. Wagai, R. & Mayer, L. M. Sorptive stabilization of organic matter in soils by hydrous iron oxides. *Geochim. Cosmochim. Acta* **71**, 25–35 (2007).
4. Hedges, J. I. & Keil, R. G. Sedimentary organic matter preservation: an assessment and speculative synthesis. *Mar. Chem.* **49**, 81–115 (1995).
5. Berner, R. A. The long-term carbon cycle, fossil fuels and atmospheric composition. *Nature* **426**, 323–326 (2003).
6. Berner, R. A. Sedimentary pyrite formation. *Am. J. Sci.* **268**, 1–23 (1970).
7. Boudot, J.-P., Bel Hadj, B. A., Steiman, R. & Seigle-Murandi, F. Biodegradation of synthetic organo-metallic complexes of iron and aluminum with selected metal to carbon ratios. *Soil Biol. Biochem.* **21**, 961–966 (1989).
8. Jones, D. L. & Edwards, A. C. Influence of sorption on the biological utilization of two simple carbon substrates. *Soil Biol. Biochem.* **30**, 1895–1902 (1998).
9. Poulton, S. W. & Raiswell, R. Chemical and physical characteristics of iron oxides in riverine and glacial meltwater sediments. *Chem. Geol.* **218**, 203–221 (2005).
10. van der Zee, C., Roberts, D. R., Rancourt, D. G. & Slomp, C. P. Nano-goethite is the dominant reactive oxyhydroxide phase in lake and marine sediments. *Geology* **31**, 993–996 (2003).
11. Canfield, D. E. The geochemistry of river particulates from the continental USA: Major elements. *Geochim. Cosmochim. Acta* **61**, 3349–3365 (1997).
12. Burdige, D. J. *Geochemistry of Marine Sediments* 107–108 (Princeton Univ. Press, 2006).
13. Schwertmann, U. Inhibitory effect of soil organic matter on the crystallization of amorphous hydroxide. *Nature* **212**, 645–646 (1966).
14. Haese, R. R. et al. Iron species determination to investigate early diagenetic reactivity in marine sediments. *Geochim. Cosmochim. Acta* **61**, 63–72 (1997).
15. Mehra, O. P. & Jackson, M. L. Iron oxide removal from soils and clays by a dithionite-citrate system buffered with sodium bicarbonate. *Clays Clay Miner.* **7**, 317–327 (1958).
16. Mayer, L. M. Extent of coverage of mineral surfaces by organic matter in marine sediments. *Geochim. Cosmochim. Acta* **63**, 207–215 (1999).
17. Ransom, B., Bennet, R. H., Baerwald, R. & Shea, K. TEM study of in situ organic matter on continental margins: occurrence and the ‘monolayer’ hypothesis. *Mar. Geol.* **138**, 1–9 (1997).
18. Mackey, D. J. & Zirino, A. Comments on trace metal speciation in seawater or do ‘onions’ grow in the sea? *Anal. Chim. Acta* **284**, 635–647 (1994).
19. Guggenberger, G. & Kaiser, K. Dissolved organic matter in soil: challenging the paradigm of sorptive preservation. *Geoderma* **113**, 293–310 (2003).
20. Baas Becking, L. G. M. & Moore, D. The relation between iron and organic matter in sediments. *J. Sedim. Petrol.* **29**, 454–458 (1959).
21. Defandre, B., Mucci, A., Gagné, J.-P., Guignard, C. & Sundby, B. Early diagenetic processes in coastal marine sediments disturbed by a catastrophic sedimentation event. *Geochim. Cosmochim. Acta* **66**, 2547–2558 (2002).
22. Rue, E. L. & Bruland, K. W. Complexation of iron(III) by natural organic ligands in the Central North Pacific as determined by a new competitive ligand equilibration/adsorptive cathodic stripping voltammetric method. *Mar. Chem.* **50**, 117–138 (1995).
23. Kraemer, S. M. Iron oxide dissolution and solubility in the presence of siderophores. *Aquat. Sci.* **66**, 3–18 (2004).
24. Arnarson, T. S. & Keil, R. G. Changes in organic matter-mineral interactions for marine sediments with varying oxygen exposure time. *Geochim. Cosmochim. Acta* **71**, 3545–3556 (2007).
25. Wang, Z.-C., Druffel, E. R. M., Griffin, S., Lee, C. & Kashgarian, M. Radiocarbon studies of organic compound classes in plankton and sediment of the Northeastern Pacific Ocean. *Geochim. Cosmochim. Acta* **62**, 1365–1378 (1998).
26. Hedges, J. I., Keil, R. G. & Benner, R. What happens to terrestrial organic matter in the ocean? *Org. Geochem.* **27**, 195–212 (1997).
27. Mayer, L. M. Surface area control on organic carbon accumulation in continental margin sediments. *Geochim. Cosmochim. Acta* **58**, 1271–1284 (1994).
28. Mayer, L. M. Speculative organic matter preservation: an assessment and speculative synthesis – a comment. *Mar. Chem.* **49**, 123–126 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work is dedicated to the memory of J. I. Hedges; in addition to being an inspiration to Y.G., he provided several sediment samples used in this work. We thank H. T. Yan for surface area measurements and the captains and crews of RV *Coriolis II* for their help during sampling missions on the St Lawrence estuary. L. N. Barazzuol is also acknowledged for her work during the first phase of the project. This work was supported by grants (Y.G. and A.M.) and scholarships (K.L.) from NSERC, CFI and FQRNT. This is GEOTOP contribution no. 2012-0001.

Author Contributions The original hypothesis was formulated by Y.G., and K.L., Y.G. and A.M. designed the project, interpreted the data and wrote the manuscript. K.L. gathered all the data. Groundwork for this study was carried out by A.O., who also contributed to the writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Y.G. (ygelinas@alcor.concordia.ca).

Experimental verification of Landauer's principle linking information and thermodynamics

Antoine Bérut¹, Artak Arakelyan¹, Artyom Petrosyan¹, Sergio Ciliberto¹, Raoul Dillenschneider² & Eric Lutz^{3†}

In 1961, Rolf Landauer argued that the erasure of information is a dissipative process¹. A minimal quantity of heat, proportional to the thermal energy and called the Landauer bound, is necessarily produced when a classical bit of information is deleted. A direct consequence of this logically irreversible transformation is that the entropy of the environment increases by a finite amount. Despite its fundamental importance for information theory and computer science^{2–5}, the erasure principle has not been verified experimentally so far, the main obstacle being the difficulty of doing single-particle experiments in the low-dissipation regime. Here we experimentally show the existence of the Landauer bound in a generic model of a one-bit memory. Using a system of a single colloidal particle trapped in a modulated double-well potential, we establish that the mean dissipated heat saturates at the Landauer bound in the limit of long erasure cycles. This result demonstrates the intimate link between information theory and thermodynamics. It further highlights the ultimate physical limit of irreversible computation.

The idea of a connection between information and thermodynamics can be traced back to Maxwell's 'demon'^{6–8}. The demon is an intelligent creature able to monitor individual molecules of a gas contained in two neighbouring chambers initially at the same temperature. Some of the molecules will be going faster than average and some will be going slower. By opening and closing a molecular-sized trap door in the partitioning wall, the demon collects the faster (hot) molecules in one of the chambers and the slower (cold) ones in the other. The temperature difference thus created can be used to run a heat engine, and produce useful work. By converting information (about the position and velocity of each particle) into energy, the demon is therefore able to decrease the entropy of the system without performing any work himself, in apparent violation of the second law of thermodynamics. A simplified, one-molecule engine introduced later⁹ has been recently realized experimentally using non-equilibrium feedback manipulation of a Brownian particle¹⁰. The paradox of the apparent violation of the second law can be resolved by noting that during a full thermodynamic cycle, the memory of the demon, which is used to record the coordinates of each molecule, has to be reset to its initial state^{11,12}. Indeed, according to Landauer's principle, any logically irreversible transformation of classical information is necessarily accompanied by the dissipation of at least $kT\ln(2)$ of heat per lost bit (about 3×10^{-21} J at room temperature (300 K)), where k is the Boltzmann constant and T is the temperature.

A device is said to be logically irreversible if its input cannot be uniquely determined from its output¹³. Any Boolean function that maps several input states onto the same output state, such as AND, NAND, OR and XOR, is therefore logically irreversible. In particular, the erasure of information, the RESET TO ONE operation, is logically irreversible and leads to an entropy increase of $k\ln(2)$ per erased bit^{14–16}. This entropy cost required to reset the demon's memory to a blank state is always larger than the initial entropy reduction, thus safeguarding the second law. Landauer's principle hence seems to be a central result that not only exorcizes Maxwell's demon, but also represents the

fundamental physical limit of irreversible computation. However, its validity has been repeatedly questioned and its usefulness criticized^{17–22}. From a technological perspective, energy dissipation per logic operation in present-day silicon-based digital circuits is about a factor of 1,000 greater than the ultimate Landauer limit, but is predicted to quickly attain it within the next couple of decades^{23,24}. Moreover, thermodynamic quantities on the scale of the thermal energy kT have been measured in mesoscopic systems such as colloidal particles in driven harmonic²⁵ and non-harmonic optical traps²⁶.

To verify the erasure principle experimentally, we consider, following the original work of Landauer¹, an overdamped colloidal particle in a double-well potential as a generic model of a one-bit memory. For this, we use a custom-built vertical optical tweezer that traps a silica bead (2 μm in diameter) at the focus of a laser beam^{27,28}. We create the double-well potential by focusing the laser alternately at two different positions with a high switching rate. The exact form of the potential is determined by the laser intensity and by the distance between the two focal points (Methods). As a result, the bead experiences an average potential $U(x, t)$, whose measured form is plotted in Fig. 1 for different stages of the erasure cycle. When the barrier is high compared with the thermal energy, kT (Fig. 1a, f), the particle is trapped in one of the potential wells. By contrast, when the barrier is low (Fig. 1b), the particle can jump from one well to the other. The state of the memory is assigned the value 0 if the particle is in the left-hand well ($x < 0$) and 1 if the particle is in the right-hand well ($x > 0$). The memory is said to be erased when its state is reset to 1 (or alternatively 0) irrespective of its initial state.

In our experiment, we follow a procedure which is quite similar to that discussed in detail in ref. 12. We start with the theoretical configuration in which the two wells are occupied with an equal probability of one-half. The initial entropy of the system is thus $S_i = k\ln(2)$. The memory is reset to 1 by first lowering the barrier height (Fig. 1b) and then applying a tilting force that brings the particle into the right-hand well (Fig. 1c–e). Finally, the barrier is increased to its initial value (Fig. 1f). At the end of this reset operation, the information initially contained in the memory has been erased and the final entropy is zero: $S_f = 0$. Thus, the minimum entropy production of this erasure process is $k\ln(2)$. The possibility of reaching this minimum depends on the timing of the procedure. The one used in our experiment is sketched in Fig. 2a. Specifically, we lower the barrier from a height larger than $8kT$ to $2.2kT$ over a time of 1 s by decreasing the power of the laser. This time is long compared with the relaxation time of the bead. We keep the barrier low for a time τ , during which we apply a linearly increasing force of maximal amplitude F_{max} , which corresponds to the tilt of the potential. We generate this force by displacing the cell containing the single bead with respect to the laser with the help of a piezoelectric motor. We close the erasure cycle by switching off the tilt and bringing the barrier back to its original height in again 1 s (Fig. 2a). A particle initially in memory state 0 will then be brought into state 1. The total duration of the erasure protocol is $\tau_{\text{cycle}} = \tau + 2$ s. Our two free parameters are the duration of the tilt, τ , and its maximal amplitude,

¹Laboratoire de Physique, École Normale Supérieure, CNRS UMR5672 46 Allée d'Italie, 69364 Lyon, France. ²Physics Department and Research Center OPTIMAS, University of Kaiserslautern, 67663 Kaiserslautern, Germany. ³Department of Physics, University of Augsburg, 86135 Augsburg, Germany. [†]Present address: Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany.

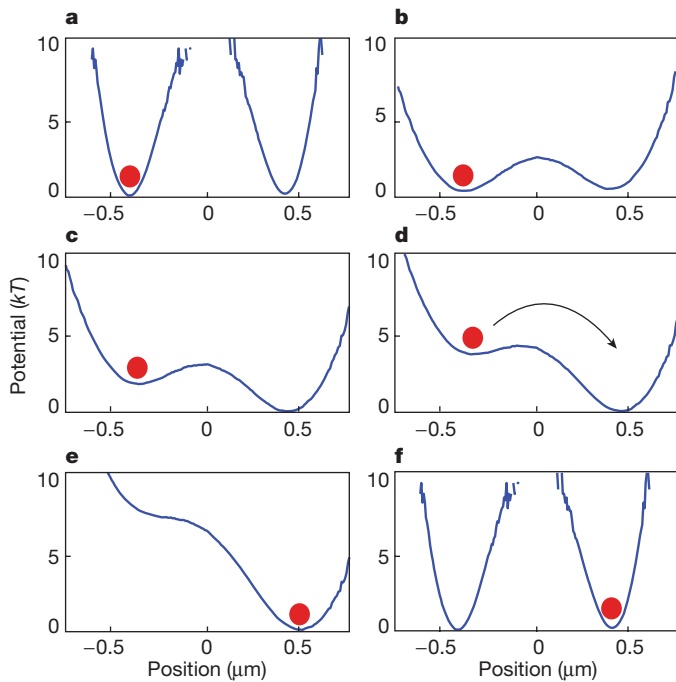


Figure 1 | The erasure protocol used in the experiment. One bit of information stored in a bistable potential is erased by first lowering the central barrier and then applying a tilting force. In the figures, we represent the transition from the initial state, 0 (left-hand well), to the final state, 1 (right-hand well). We do not show the obvious $1 \rightarrow 1$ transition. Indeed the procedure is such that irrespective of the initial state, the final state of the particle is always 1. The potential curves shown are those measured in our experiment (Methods).

F_{\max} (Methods). The second cycle in Fig. 2a corresponds to the reversed protocol, which brings the particle from state 1 to state 0 (Methods).

We use a fast camera to record the successive positions of the bead during the erasure process. A typical measured trajectory of the particle for a transition $0 \rightarrow 1$ during a cycle is shown in Fig. 2c. A trajectory for the transition $1 \rightarrow 1$ is depicted in Fig. 2d. In this case there is an instantaneous jump to the other well induced by thermal noise, but the final state is 1.

Thermodynamic quantities are stochastic variables at the microscopic level of our experiment, because thermal fluctuations cannot be neglected. The dissipated heat along a given trajectory, $x(t)$, is given by the integral²⁹ $Q = - \int_0^{\tau_{\text{cycle}}} dt \dot{x}(t) \partial U(x, t) / \partial x$. According to the laws of thermodynamics, the mean dissipated heat obtained by averaging over many trajectories is always larger than the entropy difference:

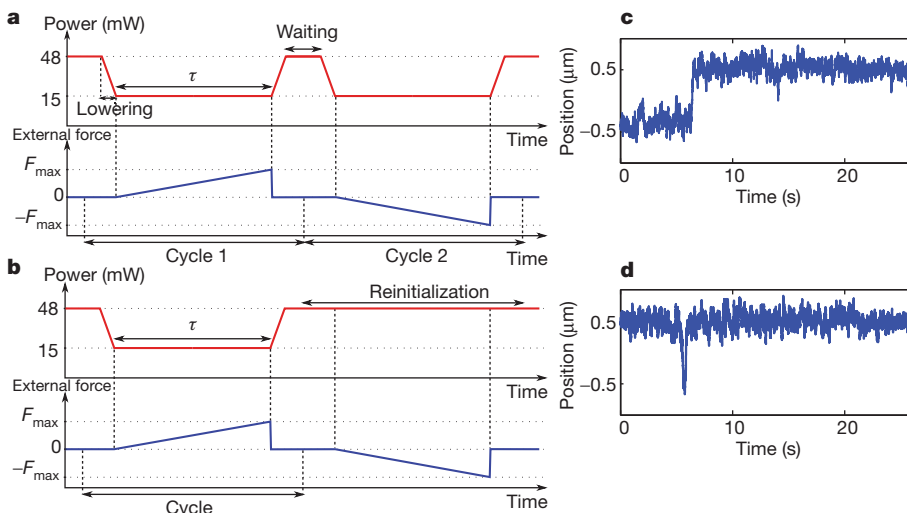


Figure 2 | Erasure cycles and typical trajectories. **a**, Protocol used for the erasure cycles bringing the bead from the left-hand well (state 0) to the right-hand well (state 1), and vice versa. **b**, Protocol used to measure the heat for the cycles in which the bead does not change wells. The reinitialization is needed to restart the measurement, but is not a part of the erasure protocol (Methods). **c**, Example of a measured bead trajectory for the transition $0 \rightarrow 1$. **d**, Example of a measured bead trajectory for the transition $1 \rightarrow 1$.

$\langle Q \rangle \geq -T\Delta S = kT \ln(2) = \langle Q \rangle_{\text{Landauer}}$. In practice, we average over situations in which the memory is either initially already in state 1 or is switched from state 0 to state 1. We typically average over more than 600 cycles. It is inconvenient to select randomly the initial configuration during two erasure cycles, so we treat the two cases independently. When the state of the memory is changed, we use a series of double cycles (Fig. 2a), which bring the bead from one well to the other, and back. In the opposite case, when the state of the memory is unmodified, we apply a series of double cycles containing a reinitialization phase (Fig. 2b). This series is useless in the erasure process itself, but is necessary to restart the measurement by keeping the bead in the initial well (Methods). We determine the dissipated heat during one erasure cycle as follows. We first note that the bead necessarily ends up in the initial state after completion of both double cycles. Because the modulation of the height of the barrier occurs on times much slower than the relaxation of the bead, it is quasi-reversible and does not contribute to the dissipated heat. We therefore only retain the contribution stemming from the external tilt, averaged over the cycles corresponding to the change of state and over the cycles in which the memory is unchanged.

A key characteristic of the erasure process is its success rate, that is, the relative number of cycles bringing the bead in the expected well. Figure 3a shows the dependence of the erasure rate on the tilt amplitude, F_{\max} . For definiteness, we have kept the product $F_{\max} \tau$ constant. We observe that the erasure rate drops sharply at low amplitudes when the tilt force is too weak to push the bead over the barrier, as expected. For large values of F_{\max} , the erasure rate saturates at around 95%. This saturation reflects the finite size of the barrier and the possible occurrence of spontaneous thermal activation into the wrong well. An example of a distribution of the dissipated heat for the transition $0 \rightarrow 1$ is displayed in Fig. 3b. Owing to thermal fluctuations, the dissipated heat may be negative and maximum erasure below the Landauer limit may be achieved for individual realizations, but not on average¹⁶.

Figure 3c shows the average dissipated heat, $\langle Q \rangle$, over a large number of erasure protocols as a function of the duration of the cycle, for various success rates. For each cycle duration, τ , we have set the amplitude, F_{\max} , of the tilt such that the erasure rate remains constant to a good approximation. For large durations, the mean dissipated heat does saturate at the Landauer limit. We observe, moreover, that incomplete erasure leads to less dissipated heat. For a success rate of r , the Landauer bound can be generalized to $\langle Q \rangle_{\text{Landauer}}^r = kT [\ln(2) + r \ln(r) + (1-r) \ln(1-r)]$. Thus, no heat needs to be produced for $r = 0.5$. In that case, the state of the memory is left unchanged by the protocol and the transformation is quasi-reversible. For ideal quasi-static erasure processes ($\tau \rightarrow \infty$), the dissipated heat is equal to the Landauer value. For large but finite τ , we can quantify the asymptotic approach to the Landauer limit by noting, following ref. 30, that $\langle Q \rangle = \langle Q \rangle_{\text{Landauer}} + B/\tau$, where B is a positive constant (Methods). For

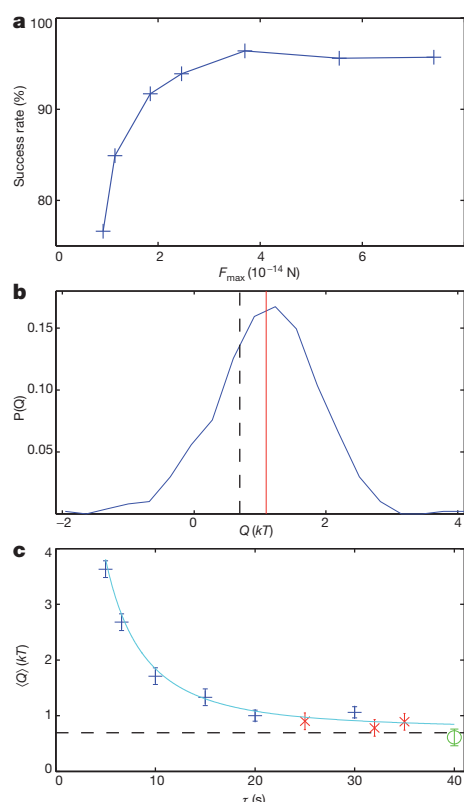


Figure 3 | Erasure rate and approach to the Landauer limit. **a**, Success rate of the erasure cycle as a function of the maximum tilt amplitude, F_{\max} , for constant $F_{\max}\tau$. **b**, Heat distribution $P(Q)$ for transition $0 \rightarrow 1$ with $\tau = 25$ s and $F_{\max} = 1.89 \times 10^{-14}$ N. The solid vertical line indicates the mean dissipated heat, $\langle Q \rangle$, and the dashed vertical line marks the Landauer limit, $\langle Q \rangle_{\text{Landauer}}$. **c**, Mean dissipated heat for an erasure cycle as a function of protocol duration, τ , measured for three different success rates, r : plus signs, $r \geq 0.90$; crosses, $r \geq 0.85$; circles, $r \geq 0.75$. The horizontal dashed line is the Landauer limit. The continuous line is the fit with the function $[A\exp(-t/\tau_K) + 1]B/\tau$, where τ_K is the Kramers time for the low barrier (Methods). Error bars, 1 s.d.

shorter durations, we find excellent agreement with an exponential relaxation, $\langle Q \rangle = \langle Q \rangle_{\text{Landauer}} + [A\exp(-t/\tau_K) + 1]B/\tau$, with a relaxation time given by the Kramers time, τ_K , for the low barrier (Methods). Our experimental results indicate that the thermodynamic limit to information erasure, the Landauer bound, can be approached in the quasi-static regime but not exceeded. They hence demonstrate one of the fundamental physical limitations of irreversible computation. Owing to the universality of thermodynamics, this limit is independent of the actual device, circuit or material used to implement the irreversible operation.

METHODS SUMMARY

We use a custom-built vertical optical tweezer made of an oil immersion objective ($\times 63$; numerical aperture, 1.4) that focuses a laser beam (wavelength, $\lambda = 1,064$ nm) to the diffraction limit for trapping glass beads^{27,28} ($2 \mu\text{m}$ in diameter). The beads are dispersed in bidistilled water at a very low concentration. The suspension is introduced in a disk-shaped cell (18 mm in diameter, 1 mm in depth), and a single bead is then trapped and moved away from the others. The position of the bead is tracked using a fast camera with a resolution of 108 nm per pixel, which after treatment gives the position with a precision greater than 10 nm. The trajectories of the bead are sampled at 502 Hz. The double-well potential is obtained by switching the laser at a rate of 10 kHz between two points separated by a distance $d_f = 1.45 \mu\text{m}$, which is kept fixed. The distance between the two minima of the double-well potential is $0.9 \mu\text{m}$. The height of the barrier is modulated by varying the power of the laser from $I_L = 48$ mW (barrier height, $> 8kT$) to $I_L = 15$ mW (barrier height, $2.2kT$). The external tilt is created by displacing the cell with respect to the laser with a piezoelectric motor, thus inducing a viscous flow. The viscous force is simply $F = -\gamma v$, where $\gamma = 1.89 \times 10^{-10}$ N s m^{-1} is the coefficient of friction and v is the velocity of the cell. In the erasure protocol, the

amplitude of the viscous force is increased linearly during time τ : $F(t) = \pm F_{\max}t/\tau$. The heat dissipated by the tilt is

$$Q = - \int_0^{\tau_{\text{cycle}}} dt F(t) \dot{x}(t) = \pm \int_0^{\tau} dt F_{\max}(t/\tau) \dot{x}(t)$$

The velocity is computed using the discretization $\dot{x}(t + \Delta t/2) \approx [x(t + \Delta t) - x(t)]/\Delta t$.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 October 2011; accepted 17 January 2012.

- Landauer, R. Irreversibility and heat generation in the computing process. *IBM J. Res. Develop.* **5**, 183–191 (1961).
- Landauer, R. Dissipation and noise immunity in computation and communication. *Nature* **335**, 779–784 (1988).
- Lloyd, S. Ultimate physical limits to computation. *Nature* **406**, 1047–1054 (2000).
- Meindl, J. D. & Davis, J. A. The fundamental limit on binary switching energy for terascale integration. *IEEE J. Solid-state Circuits* **35**, 1515–1516 (2000).
- Plenio, M. B. & Vitelli, V. The physics of forgetting: Landauer's erasure principle and information theory. *Contemp. Phys.* **42**, 25–60 (2001).
- Brillouin, L. *Science and Information Theory* (Academic, 1956).
- Leff, H. S. & Rex, A. F. *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing* (IOP, 2003).
- Maruyama, K., Nori, F. & Vedral, V. The physics of Maxwell's demon and information. *Rev. Mod. Phys.* **81**, 1–23 (2009).
- Szilard, L. On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. *Z. Phys.* **53**, 840–856 (1929).
- Toyabe, S., Sagawa, T., Ueda, M., Muneyuki, E. & Sano, M. Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality. *Nature Phys.* **6**, 988–992 (2010).
- Penrose, O. *Foundations of Statistical Mechanics: A Deductive Treatment* (Pergamon, 1970).
- Bennett, C. H. The thermodynamics of computation: a review. *Int. J. Theor. Phys.* **21**, 905–940 (1982).
- Bennett, C. H. Logical reversibility of computation. *IBM J. Res. Develop.* **17**, 525–532 (1973).
- Shizume, K. Heat generation required by information erasure. *Phys. Rev. E* **52**, 3495–3499 (1995).
- Piechocinska, P. Information erasure. *Phys. Rev. A* **61**, 062314 (2000).
- Dillenschneider, R. & Lutz, E. Memory erasure in small systems. *Phys. Rev. Lett.* **102**, 210601 (2009).
- Earman, J. & Norton, J. D. EXORCIST XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond. *Stud. Hist. Phil. Sci. B* **30**, 1–40 (1999).
- Shenker, O. R. Logic and entropy. Preprint at (<http://philsci-archive.pitt.edu/115/>) (2000).
- Maroney, O. J. E. The (absence of a) relationship between thermodynamic and logical reversibility. *Studies Hist. Phil. Sci. B* **36**, 355–374 (2005).
- Norton, J. D. Eaters of the lotus: Landauer's principle and the return of Maxwell's demon. *Stud. Hist. Phil. Sci. B* **36**, 375–411 (2005).
- Sagawa, T. & Ueda, M. Minimal energy cost for thermodynamic information processing: measurement and information erasure. *Phys. Rev. Lett.* **102**, 250602 (2009).
- Norton, J. D. Waiting for Landauer. *Stud. Hist. Phil. Sci. B* **42**, 184–198 (2011).
- Frank, M. P. The physical limits of computing. *Comput. Sci. Eng.* **4**, 16–26 (2002).
- Pop, E. Energy dissipation and transport in nanoscale devices. *Nano Res.* **3**, 147–169 (2010).
- Wang, G. M., Sevick, E. M., Mittag, E., Searles, D. J. & Evans, D. J. Experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales. *Phys. Rev. Lett.* **89**, 050601 (2002).
- Blickle, V., Speck, T., Helden, L., Seifert, U. & Bechinger, C. Thermodynamics of a colloidal particle in a time-dependent nonharmonic potential. *Phys. Rev. Lett.* **96**, 070603 (2006).
- Jop, P., Petrosyan, A. & Ciliberto, S. Work and dissipation fluctuations near the stochastic resonance of a colloidal particle. *Europhys. Lett.* **81**, 50005 (2008).
- Gomez-Solano, J. R., Petrosyan, A., Ciliberto, S., Chetrite, R. & Gawedzki, K. Experimental verification of a modified fluctuation-dissipation relation for a micron-sized particle in a nonequilibrium steady state. *Phys. Rev. Lett.* **103**, 040601 (2009).
- Sekimoto, K. *Stochastic Energetics* (Springer, 2010).
- Sekimoto, K. & Sasa, S. I. Complementarity relation for irreversible process derived from stochastic energetics. *J. Phys. Soc. Jpn* **6**, 3326–3328 (1997).

Acknowledgements This work was supported by the Emmy Noether Program of the DFG (contract no. LU1382/1-1), the Cluster of Excellence Nanosystems Initiative Munich (NIM), DAAD, and the Research Center Transregio 49 of the DFG.

Author contributions All authors contributed substantially to this work.

Author information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to E.L. (eric.lutz@fu-berlin.de).

METHODS

Experimental set-up. We use a custom-built vertical optical tweezer made of an oil immersion objective ($\times 63$; numerical aperture, 1.4) that focuses a laser beam (wavelength, $\lambda = 1,064$ nm) to the diffraction limit for trapping glass beads^{27,28} ($2\ \mu\text{m}$ in diameter). The beads are dispersed in bidistilled water at a very low concentration. The suspension is introduced in a disk-shaped cell (18 mm in diameter, 1 mm in depth), and a single bead is then trapped and moved away from the others. This step is important to stop the trapped bead being perturbed by other Brownian particles during the measurement. The position of the bead is tracked using a fast camera with a resolution of 108 nm per pixel, which after treatment gives the position with a precision greater than 10 nm. The trajectories of the bead are sampled at 502 Hz. The double-well potential is obtained by switching the laser at a rate of 10 kHz between two points by a distance $d_f = 1.45\ \mu\text{m}$, which is kept fixed. The form of the potential, which is a function of d_f and the laser intensity, I_L , can be determined in equilibrium by measuring the probability, $P(x, I_L) = \text{Nexp}[-U_0(x, I_L)/kT]$, of the bead being at position x , which implies that $U_0(x, I_L) = -kT \ln[P(x, I_L)/N]$ (Fig. 1a, b, f). The distribution $P(x, I_L)$ is estimated from about 10^6 samples. The measured $U_0(x, I_L)$ are plotted in Fig. 1a, b, f and can be fitted by an eighth-order polynomial: $U_0(x, I_L) = \sum_{n=0}^8 u_n(I_L, d_f) x^n$. The distance between the two minima of the double-well potential is $0.9\ \mu\text{m}$. The two wells are nearly symmetrical, with a maximum energy difference of $0.4kT$. The height of the barrier, ΔU , is modulated by varying the power of the laser from $I_L = 48$ mW (barrier height, $>8kT$) to $I_L = 15$ mW (barrier height, $2.2kT$). In equilibrium for a barrier of $8kT$, the characteristic jumping time (Kramers time), $\tau_K = \tau_0 \exp(\Delta U/kT)$, between the two wells is about $3,000$ s, which is much longer than any timescale in the experiment ($\tau_0 \approx 1$ s in our experiment).

The external tilt is created by displacing the cell with respect to the laser with a piezoelectric motor, thus inducing a viscous flow. The viscous force is simply $F = -\gamma v$, where $\gamma = 1.89 \times 10^{-10} \text{ N s m}^{-1}$ is the coefficient of friction and v is the velocity of the cell. In the erasure protocol, the amplitude of the viscous force is increased linearly during time τ : $F(t) = \pm F_{\text{max}} t/\tau$. In Fig. 1c–e, we plot $U(x, t) = U_0(x, I_L) - F(t)x$ for $I_L = 15$ mW and for three different values of t . The reinitialization procedure shown in Fig. 2b is necessary to displace the cell to its initial position, but it does not contribute to the erasure process. We note that, unlike the useful erasure cycles, this reinitialization is performed when the barrier is high. Thus, the bead remains always in the same well.

Heat measurements. The heat dissipated by the tilt is

$$Q = - \int_0^{\tau_{\text{cycle}}} dt F(t) \dot{x}(t) = \pm \int_0^{\tau} dt F_{\text{max}}(t/\tau) \dot{x}(t)$$

The velocity is computed using the discretization $\dot{x}(t + \Delta t/2) \approx [x(t + \Delta t) - x(t)]/\Delta t$. To characterize the asymptotic approach to the Landauer bound we use the fact that, in the quasi-static limit ($\tau \rightarrow \infty$), the mean work, $\langle W \rangle$, can be expressed in terms of the free energy difference, ΔF , as $\langle W \rangle \approx \Delta F + B/\tau$ (ref. 30). According to the first law of thermodynamics, $\langle \Delta U \rangle = \langle W \rangle - \langle Q \rangle = 0$ for a cycle, and we therefore find that $\Delta F = -T\Delta S$ and $\langle Q \rangle = \langle W \rangle \approx kT \ln(2) + B/\tau$. This asymptotic result is generic and does not depend on the details of the potential. For shorter times, we find an exponential relaxation of the form $\langle Q \rangle = kT \ln(2) + [A \exp(-t/\tau_K) + 1]B/\tau$, where τ_K is the Kramers time for the low barrier. The presence of the characteristic Kramers time can be understood by noting that cycles that last more than a few multiples of τ_K are very efficient for erasure because the probability that a jump to the right-hand well occurs by thermal activation is greatly increased.

Field-driven photoemission from nanostructures quenches the quiver motion

G. Herink¹, D. R. Solli¹, M. Gulde¹ & C. Ropers¹

Strong-field physics, an extreme limit of light–matter interaction^{1–3}, is expanding into the realm of surfaces^{4,5} and nanostructures^{6–11} from its origin in atomic and molecular science^{12–15}. The attraction of nanostructures lies in two intimately connected features: local intensity enhancement and sub-wavelength confinement of optical fields. Local intensity enhancement facilitates access to the strong-field regime and has already sparked various applications, whereas spatial localization has the potential to generate strong-field dynamics exclusive to nanostructures. However, the observation of features unattainable in gaseous media is challenged by many-body effects and material damage, which arise under intense illumination of dense systems^{16–19}. Here, we non-destructively access this regime in the solid state by employing single plasmonic nanotips and few-cycle mid-infrared pulses, making use of the wavelength-dependence of the interaction, that is, the ponderomotive energy. We investigate strong-field photoelectron emission and acceleration from single nanostructures over a broad spectral range, and find kinetic energies of hundreds of electronvolts. We observe the transition to a new regime in strong-field dynamics, in which the electrons escape the nanolocalized field within a fraction of an optical half-cycle. The transition into this regime, characterized by a spatial adiabaticity parameter, would require relativistic electrons in the absence of nanostructures. These results establish new degrees of freedom for the manipulation and control of electron dynamics on femtosecond and attosecond timescales, combining optical near-fields and nanoscopic sources.

The explanation of the photoelectric effect by Einstein in 1905 marks a defining moment in the turn towards modern physics, reconciling numerous observations contradicting classical expectations²⁰. Einstein's contemporaries had been puzzled by the scaling of photoelectron kinetic energies: increasing with light frequency and independent of light intensity. Some quotations from Lenard's 1902 experimental paper illustrate the historical confusion. He expected that “the part of the initial velocity stemming from the light...should have been acquired within the last half or full resonance oscillation, hence it ought to grow with light intensity,” but in experiments found “not the slightest dependency...on light intensity”²¹. Here we show that nanostructures exposed to intense mid-infrared fields may indeed cause classical electron dynamics determined by less than an optical half-cycle.

Today, it is known that strong-field interactions of light with atoms and surfaces, particularly electron emission, are governed by both quantum-mechanical and classical concepts, given that the final state of a photoemission event is influenced by the presence of the optical field¹. The underlying principles have been widely studied in atoms and molecules^{2,3,12–15,22} and classical scalings become increasingly pronounced at higher intensity and longer wavelength^{3,15}. The photon-driven (quantum) and the field-driven (classical) regimes are regarded as limiting cases¹ with a transition characterized by the Keldysh parameter $\gamma = \omega/\omega_t$. Also known as the adiabaticity parameter, it relates the optical driving frequency ω to a characteristic tunnelling frequency $\omega_t = eF/\sqrt{2m\Phi}$ (where F is the electric field, Φ is the work

function, m is the mass of the electron and e is its charge). In the strong-field/low-frequency limit ($\gamma \ll 1$), the emission is viewed as tunnelling that adiabatically follows the momentary optical field. The energies of photoelectrons are then determined by their quiver motion in the driving field^{2,3} and their energy cutoffs scale with the ponderomotive potential $U_p = e^2 F^2 / 4m\omega^2$. These two steps, adiabatic tunnelling and propagation of classical point particles, form the basis of the so-called Simpleman model^{2,3}, which accurately describes numerous aspects of strong-field atomic ionization.

The standard Simpleman description of electron trajectories contains only the temporal—not the spatial—dependence of the driving field because the wavelengths or typical optical focus diameters are usually much larger than the quiver amplitude. Simple considerations arising from the diffraction limit show that this condition generally holds for non-relativistic energies. In contrast, the optical fields at nanostructures may vary greatly on such scales, so that entirely different electron dynamics, which have not yet been observed, may be expected (see sketches in Fig. 1a, b). Recently, numerous studies have addressed localized emission of electrons from nanometric metallic needles using 800-nm excitation radiation^{6–8,23–25}. The possibilities for local enhancement and control of optical fields make strong-field physics in nanostructures highly attractive^{6–11,26,27}. However, many-body effects complicate matters, and short-wavelength damage thresholds limit Keldysh parameters close to unity, with ponderomotive acceleration just arising with energies around 10 eV. Higher energies have been achieved in plasmonic fields at planar surfaces^{4,5},

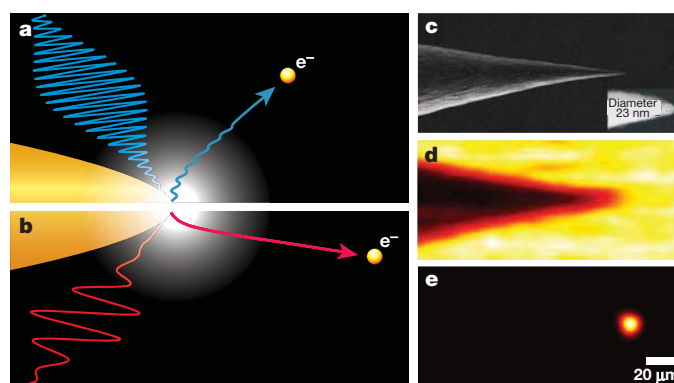


Figure 1 | Electron emission from nanotips using ultrashort mid-infrared fields. Trajectories of photoelectrons generated by intense light fields depend strongly on whether the quiver amplitude is smaller (a, short-wavelength excitation) or larger (b, long-wavelength excitation) than the characteristic decay length of the optical near-field (bright white region). In b, the electron escapes the near-field before substantial back acceleration. c, Scanning electron micrograph of gold tip used in measurements. Inset, apex region and tip diameter. d and e, Mid-infrared shadow image (d) and electron-emission image (e), obtained by scanning the tip through the focal plane, demonstrating exclusive electron emission from the tip apex (wavelength 3.8 μm). We note that the spot size in e represents the focus diameter, while the actual emission region has nanometre dimensions.

¹Courant Research Center for Nano-Spectroscopy and X-Ray Imaging, University of Göttingen, 37077 Göttingen, Germany.

but interpretations can be complicated by inhomogeneous broadenings from surface roughness and the inclusion of space charge^{16–19}.

Here, we study ultrafast strong-field photoemission from single nanostructures over a wide range of wavelengths up to 8 μm , reaching deep into the tunnelling regime, with Keldysh parameters as low as 0.1. We identify wavelength scalings of classical origin that differ substantially from usual ponderomotive behaviour. This follows directly from particle dynamics in fields decaying over distances shorter than the quiver amplitude. In particular, we observe a transition to a new sub-cycle regime, in which the electrons' kinetic energy, reaching hundreds of electronvolts, becomes largely wavelength-independent. This transition is characterized by introducing a spatial complement to the Keldysh parameter.

In our experiments, tunable femtosecond pulses are focused onto electrochemically etched gold tips (radius of curvature about 10 nm; Fig. 1c) in ultrahigh vacuum. The kinetic energy spectra of photoelectrons are recorded as a function of intensity across the near- and mid-infrared spectrum. Figure 1d and e illustrates the confinement of the photoemission to the apex (wavelength 3.8 μm) with simultaneous measurements of the light passing the tip (Fig. 1d) and the induced electron emission (Fig. 1e) from scanning the tip in the focal plane. We find such apex-exclusive signals at all wavelengths, evidencing the ultra-broadband nature of both field enhancements (arising from geometry and plasmonic resonance) and nonlinear photoemission at these tips^{23,28,29}. Electron spectra (Fig. 2a) from two tips of different radii and at various intensities show the scaling of kinetic energies up to cutoffs of hundreds of electronvolts.

To examine the observed behaviour, we compute electron spectra using experimental parameters in a two-step model adapted to localized fields. The simulations (Fig. 2a, solid lines) include a Fowler–Nordheim tunnelling model (justified, for example, in ref. 30),

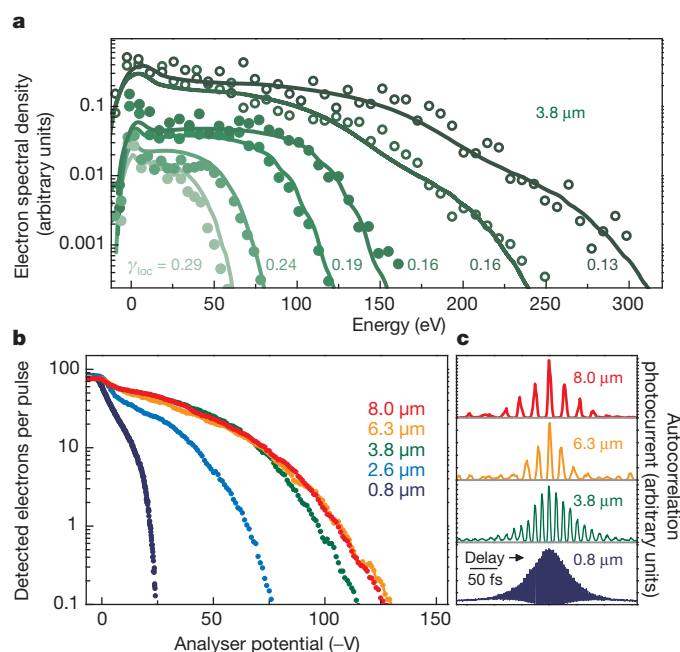


Figure 2 | Experimental photoelectron spectra and autocorrelations at mid-infrared wavelengths. **a**, Experimental (circles) and simulated (solid lines) kinetic energy distributions of photoelectrons for increasing intensities at wavelength 3.8 μm . Tip radii are 12 nm (solid circles) and 22 nm (open circles); intensity enhancements are 52 and 34, respectively. **b**, Raw data of energy analyser scans for various wavelengths and fixed total charge (12 nm tip). Curves represent number of electrons with energies above analyser potential and demonstrate wavelength-dependent cutoff. (A negative bias voltage is needed to repel the negatively charged electrons.) **c**, Interferometric autocorrelations (electron signal) of pulses at different wavelengths (charge not held fixed for different traces in contrast to **b**).

interaction of the electrons with a strongly localized tip field and rescattering at the surface (see Methods Summary). Good agreement with the experimental data is obtained. Throughout this work, we use the local (that is, tip-enhanced) fields in expressions and values of the Keldysh parameter γ_{loc} . For example, the curve extending to 300 eV in Fig. 2a ($\gamma_{\text{loc}} = 0.13$) corresponds to a local peak field strength of 28 V nm^{-1} . The two tips of different radii exhibit different total currents and cutoffs at the same $\gamma_{\text{loc}} = 0.16$ for different emission areas (influencing the total current) and field decay lengths. As we discuss later, tighter confinement in fact leads to a cutoff reduction.

At such high energies and significant electron densities, the influence of many-body effects (space charge) should be considered^{16–19}. In recent ensemble measurements on dielectric nanospheres with near-infrared light, space charge was found to contribute substantially to particle dynamics⁹, which is in contrast to the conditions of our study here. In general, emission from single sharp and moderately voltage-biased metal tips is less susceptible to space charge, in part because of rapidly diverging trajectories. Experimentally, we can verify that space charge effects are insignificant here by measuring the kinetic energy distributions at different wavelengths and constant emitted charge (Fig. 2b). The strong increase in cutoff energy with wavelength demonstrates that the dynamics are governed not by the total emitted charge, but rather by the properties of the driving field.

To identify the wavelength-scaling of the electron acceleration, conditions at fixed local driving field are desired. Given that the field enhancement generally depends on wavelength, knowledge of the incident intensity alone is insufficient. However, in the adiabatic regime, measurements of total charge and pulse duration can be used to ensure constant local fields. At constant emitted charge and the present field strengths, the pulse durations need only be approximately equal to realize nearly identical local fields; this follows from the fact that the charge in the Fowler–Nordheim model depends linearly on the pulse duration and exponentially on the inverse field. We characterize the pulse durations for different wavelengths *in situ* by overlapping a pair of identical pulses with variable delay on the tip^{23,25}, using the electron signal to produce an interferometric autocorrelation (Fig. 2c). The autocorrelations can be used to extract the pulse duration, taking into account the field-dependent tunnelling nonlinearity (Supplementary Information), which leads to the observed sharpening of the oscillatory features at lower fields, visible for example, in the 8- μm measurement. The spread in extracted pulse durations (from 60 to 95 fs) corresponds to a variation of less than $\pm 5\%$ in local field around a value of 21 V nm^{-1} for the curves shown in Fig. 2b. This allows us to study the pure wavelength-dependence of the electron dynamics.

From the scaling of the ponderomotive potential U_p , one may expect the cutoff energies for constant local intensity to depend quadratically on wavelength. We observe a drastic deviation from this behaviour. Instead, we find a strong increase of the kinetic energy for short wavelengths only, followed by a much slower increase, shown for two different local intensities in Fig. 3a. The simulations reveal that the observed deviation from regular ponderomotive scaling arises directly from the sub-wavelength decay of the near-field. The data fall into ‘corridors’ (grey-shaded areas) given by the cutoffs for zero and unity reflection coefficient at the surface. The latter is used in Fig. 2a, as in ref. 7, but we note that the finite uncertainty in the field enhancement does not allow for a precise determination of the actual reflection coefficients from the data. This might be surprising, because a factor of at least five is expected between the direct and rescattered cutoffs³. However, the cutoffs converge at longer wavelengths, a feature unique to strong-field dynamics at nanostructures (see below).

At the experimental near-field intensities and especially at the long wavelengths used, the quiver amplitude can easily exceed the nanometre-scale field decay length. If this happens, the quiver motion is increasingly quenched and the cutoff energy begins to saturate, as predicted by the simulations (Fig. 3a, solid and dashed lines). The electron propagation

in this localized field suggests the introduction of a new dimensionless adiabaticity parameter that relates the $(1/e)$ -decay length l_F of the optical field to the electron quiver amplitude $l_q = eF/m\omega^2$ as follows:

$$\delta = \frac{l_F}{l_q} = \frac{l_F m \omega^2}{eF}$$

Whereas the Keldysh parameter describes the degree of adiabaticity of the emission process (that is, in step 1 of the Simpleman model), the ratio δ characterizes the propagation and acceleration (step 2 in the Simpleman model). For $\delta \gg 1$, the electron quivers in a nearly homogeneous field over multiple optical cycles (depending on the light pulse duration) in the usual way. For $\delta \ll 1$, electrons rapidly escape from the tip-enhanced field within a time much shorter than an optical cycle (Fig. 3c, d). In fact, the same expression for the δ -parameter follows from the ratio of the escape time at maximum energy ($2U_p$) to the optical period, or similarly from the ratio of an electrostatic energy Fl_F to the ponderomotive energy. In the sub-cycle regime, the electron energy begins to scale sublinearly with intensity and will become linear in the field at even lower driving frequencies.

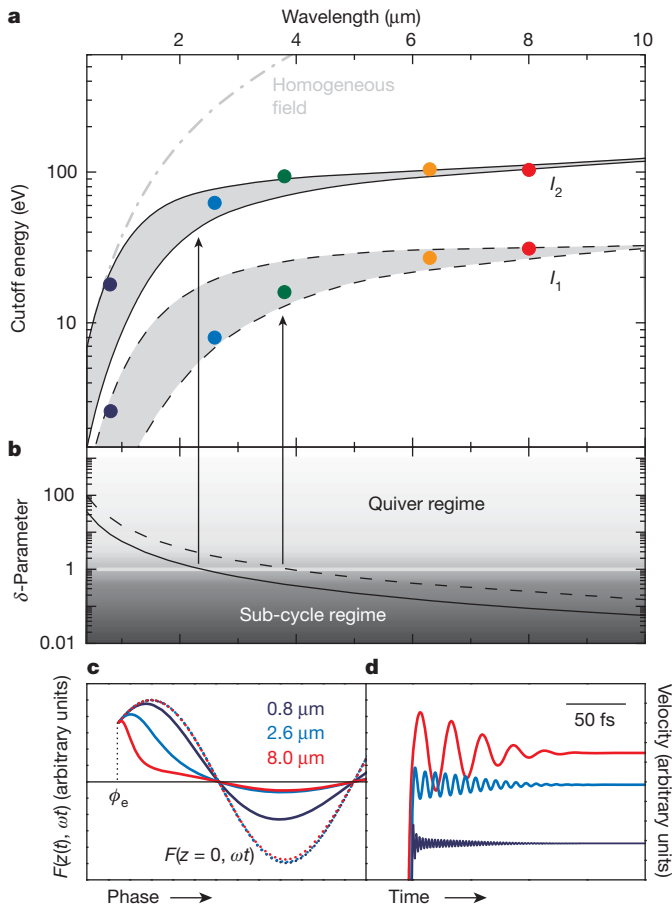


Figure 3 | Transition from quiver to sub-cycle electron dynamics.

a, Experimental cutoff energies versus wavelength for two values of emitted charge (circles). In the adiabatic limit, constant charge and constant intensity are equivalent. The cutoff is defined to include 99% of the electron population. Simulations (at local intensities $I_1 = 5.4 \text{ TW cm}^{-2}$ and $I_2 = 40 \text{ TW cm}^{-2}$) are shown for full rescattering and without any rescattering (upper and lower bounds of the grey ‘corridors’, respectively), and neglecting the spatial field dependence (dash-dotted grey line). **b**, Spatial adiabaticity parameter and transition between quiver and sub-cycle regimes corresponding to **a**. **c**, Electric field versus phase experienced by propagating electron (emitted at a phase ϕ_e) for different wavelengths; $z(t)$ is the distance from the tip. The dotted trace is the field at the surface ($z = 0$). At long wavelengths, the electron leaves the near-field within a fraction of a half-cycle. **d**, Corresponding time-dependent electron velocities. For $8 \mu\text{m}$, the electron momentum is attained in much less than one half cycle.

The transition from quiver to sub-cycle dynamics ($\delta \approx 1$) occurs within the range of wavelengths used in our experiments, as shown in Fig. 3a, b. Near the transition, the cutoff energy becomes wavelength-independent, in agreement with simulation, because the fastest electrons escape the tip-enhanced field well before it completes a half-cycle. Such conditions are only reached in the presence of both high fields and local confinement, as shown by simulations without spatial field dependence.

By means of simulations with fixed (enhanced) surface intensity and variable decay length, we find that several aspects of the electron dynamics change substantially from the quiver to the sub-cycle situation. In the sub-cycle regime, most trajectories display minimal quiver motion, and rescattering occurs over a reduced range of emission phases (Fig. 4a, b). Also, the maximum of kinetic energy shifts to later emission phases (Supplementary Information). From the usual maximum-energy values³ of $2U_p$ and $10U_p$ (defined in terms of local, not incident intensity) at long field decay length, the energies of direct and rescattered electrons become smaller and converge for $\delta \leq 1$ (Fig. 4c). This follows from reduced back-acceleration after the first half-cycle, which also causes an energy maximum at intermediate decay lengths for direct trajectories. In the present experiments, the carrier-envelope phase of the pulses is not stabilized. Recent work showed the importance of carrier-envelope phase effects, especially for the high-energy cutoff region of the spectra^{7,9}. Our simulations predict similar features also in the sub-cycle regime studied here, with modifications arising from the fact that the phase of maximum electron energy is shifted for small δ -parameters owing to the field localization (Supplementary Information).

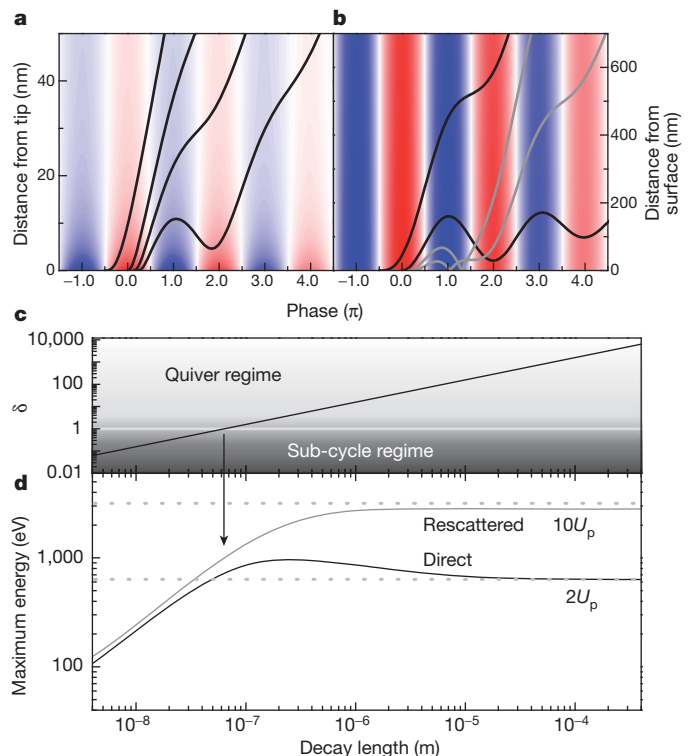


Figure 4 | Impact of field localization on electron dynamics. **a** and **b**, Simulated electron trajectories for four emission phases in localized (**a**) and homogeneous (**b**) pulsed fields (wavelength $8 \mu\text{m}$, colour shading indicates field; red and blue indicate positive and negative electric force on electron, respectively). Grey lines are rescattered trajectories. **c** and **d**, Spatial adiabaticity parameter (**c**) and maximum kinetic energies (**d**) for direct (black line) and rescattered (grey line) trajectories as a function of field decay length for a fixed peak intensity at the surface. Direct and rescattered energies converge at small δ -parameters. U_p includes the local intensity enhancement, held constant for all decay lengths in this figure for illustrative purposes.

In conclusion, a new regime in strong-field physics is reached by driving the optical near-field in nanostructures with few-cycle mid-infrared pulses. Qualitatively different electron dynamics, for example, regarding the roles of ponderomotive acceleration and backscattering, result from quiver amplitudes that exceed the near-field decay length. Here, electrons escape the driving field within a fraction of the driving period—conditions which would require relativistic energies inside diffraction-limited far-fields. In the extreme limit, the sub-cycle regime encodes the instantaneous light field, rather than its temporal integral, onto the electron energy. This will enable new applications in a strong-field ‘lab on a tip’ with ultrafast streaking and gating at large bandwidth and additional control parameters, such as the static field. The locally accelerated electrons have excellent spatial coherence and may be used as local probes of fields and atomic structures, and in energy-gated schemes, as switchable high-brightness wavepackets for time-resolved surface diffraction.

METHODS SUMMARY

A titanium-sapphire amplifier (800 nm, 1 kHz) pumps an optical parametric amplifier. From the signal and idler, tunable mid-infrared pulses (2.6–8 μm) are produced by difference frequency generation. After variable attenuation, the pulses are coupled into an ultrahigh-vacuum chamber and focused by a parabolic mirror onto the nanotip. Experiments using attenuated 800-nm pulses are also performed. Raster-scanned shadow images (for example, Fig. 1d) are recorded using the light passing the tip. Autocorrelations of the pulses are obtained with an ultra-broadband interferometer, making use of the nonlinearity of the photoemission process.

Gold nanotips are produced by electrochemical etching, achieving apex radii of 10–20 nm. Each tip is mounted on a three-dimensional micro-translation stage, allowing for raster scans and alignment in the focus. A bias voltage of -40 V is applied to the tip to draw the photoelectrons moderately away. Electrons are detected with a microchannel-plate phosphor-screen (MCP-PS) assembly. Kinetic energy distributions are measured using a retarding-field energy analyser, correcting for the static offset. A charge-coupled device (CCD) camera records the signal from the MCP-PS as described in ref. 6. The Supplementary Information contains a sketch of the set-up.

Electron spectra are simulated with a two-step model: tunnel emission of photoelectrons and subsequent acceleration in the localized, tip-enhanced laser field. Single-particle trajectories are evaluated as a function of emission phase by solving the equation of motion. The results are weighted by the appropriate tunnelling currents to produce the energy distribution. Spectra are averaged over different carrier-envelope phases. Experimental parameters (wavelength, pulse duration, pulse energy, apex and focus diameters) and intensity enhancements consistent with previous literature are used for different tips. More details are in the Supplementary Information.

Received 13 October 2011; accepted 17 January 2012.

- Keldysh, L. V. Ionization in the field of a strong electromagnetic wave. *Sov. Phys. JETP* **20**, 1307–1314 (1965).
- Corkum, P. B. Plasma perspective on strong-field ionization. *Phys. Rev. Lett.* **71**, 1994–1997 (1993).
- Paulus, G. G., Becker, W. & Walther, H. Classical rescattering effects in two-color above-threshold ionization. *Phys. Rev. A* **52**, 4043–4053 (1995).
- Irvine, S. E., Dechant, A. & Elezzabi, A. Y. Generation of 0.4-keV femtosecond electron pulses using impulsively excited surface plasmons. *Phys. Rev. Lett.* **93**, 184801 (2004).
- Rác, P. *et al.* Strong-field plasmonic electron acceleration with few-cycle, phase-stabilized laser pulses. *Appl. Phys. Lett.* **98**, 111116 (2011).
- Bormann, R., Gulde, M., Weismann, A., Yalunin, S. V. & Ropers, C. Tip-enhanced strong-field photoemission. *Phys. Rev. Lett.* **105**, 147601 (2010).
- Krüger, M., Schenk, M. & Hommelhoff, P. Attosecond control of electrons emitted from a nanoscale metal tip. *Nature* **475**, 78–81 (2011).
- Schenk, M., Krüger, M. & Hommelhoff, P. Strong-field above-threshold photoemission from sharp metal tips. *Phys. Rev. Lett.* **105**, 257601 (2010).
- Zherebtsov, S. *et al.* Controlled near-field enhanced electron acceleration from dielectric nanospheres with intense few-cycle laser fields. *Nature Phys.* **7**, 656–662 (2011).
- Kim, S. *et al.* High-harmonic generation by resonant plasmon field enhancement. *Nature* **453**, 757–760 (2008).
- Sivis, M., Duwe, M., Abel, B. & Ropers, C. Nanostructure-enhanced atomic line emission. <http://dx.doi.org/10.1038/nature10978> *Nature* (in the press).
- Hentschel, M. *et al.* Attosecond metrology. *Nature* **414**, 509–513 (2001).
- Krausz, F. & Ivanov, M. Attosecond physics. *Rev. Mod. Phys.* **81**, 163–234 (2009).
- Sansone, G. *et al.* Isolated single-cycle attosecond pulses. *Science* **314**, 443–446 (2006).
- Colosimo, P. *et al.* Scaling strong-field interactions towards the classical limit. *Nature Phys.* **4**, 386–389 (2008).
- Gilton, T. L., Cowin, J. P., Kubiak, G. D. & Hamza, A. V. Intense surface photoemission: space charge effects and self-acceleration. *J. Appl. Phys.* **68**, 4802–4810 (1990).
- Aeschlimann, M. *et al.* Observation of surface enhanced multiphoton photoemission from metal surfaces in the short pulse limit. *J. Chem. Phys.* **102**, 8606–8613 (1995).
- Petite, G., Agostini, P., Trainham, R., Mevel, E. & Martin, P. Origin of the high-energy electron emission from metals under laser irradiation. *Phys. Rev. B* **45**, 12210–12217 (1992).
- Passlack, S. *et al.* Space charge effects in photoemission with a low repetition, high intensity femtosecond laser source. *J. Appl. Phys.* **100**, 024912 (2006).
- Einstein, A. Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Ann. Phys.* **322**, 132–148 (1905).
- Lenard, P. Ueber die lichtelektrische Wirkung. *Ann. Phys.* [transl. C.R. and G.H.] **313**, 149–198 (1902).
- Meckel, M. *et al.* Laser-induced electron tunneling and diffraction. *Science* **320**, 1478–1482 (2008).
- Ropers, C., Solli, D. R., Schulz, C. P., Lienau, C. & Elsaesser, T. Localized multiphoton emission of femtosecond electron pulses from metal nanotips. *Phys. Rev. Lett.* **98**, 043907 (2007).
- Hommelhoff, P., Sortais, Y., Aghajani-Talesh, A. & Kasevich, M. A. Field emission tip as a nanometer source of free electron femtosecond pulses. *Phys. Rev. Lett.* **96**, 077401 (2006).
- Hommelhoff, P., Kealhofer, C. & Kasevich, M. A. Ultrafast electron pulses from a tungsten tip triggered by low-power femtosecond laser pulses. *Phys. Rev. Lett.* **97**, 247402 (2006).
- Süßmann, F. & Kling, M. F. Attosecond nanoplasmonic streaking of localized fields near metal nanospheres. *Phys. Rev. B* **84**, 121406 (2011).
- Stockman, M. I., Kling, M. F., Kleineberg, U. & Krausz, F. Attosecond nanoplasmonic-field microscope. *Nature Photon.* **1**, 539–544 (2007).
- Bouhelier, A., Beversluis, M., Hartschuh, A. & Novotny, L. Near-field second-harmonic generation induced by local field enhancement. *Phys. Rev. Lett.* **90**, 013903 (2003).
- Huber, A. J., Keilmann, F., Wittborn, J., Aizpurua, J. & Hillenbrand, R. Terahertz near-field nanoscopy of mobile carriers in single semiconductor nanodevices. *Nano Lett.* **8**, 3766–3770 (2008).
- Yalunin, S. V., Gulde, M. & Ropers, C. Strong-field photoemission from surfaces: theoretical approaches. *Phys. Rev. B* **84**, 195426 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Bormann, F. Schenk, M. Sivis and S. V. Yalunin for discussions. Financial support by the Deutsche Forschungsgemeinschaft (DFG-ZUK 45/1 and SPP 1391) is gratefully acknowledged.

Author Contributions All authors were closely involved in this study and contributed to the ideas, realization of the experiments, data analysis and interpretation, and writing of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.R. (croppers@gwdg.de).

CORRECTIONS & AMENDMENTS

CORRIGENDUM

doi:10.1038/nature10943

Corrigendum: The Amazon basin in transition

Eric A. Davidson, Alessandro C. de Araújo, Paulo Artaxo, Jennifer K. Balch, I. Foster Brown, Mercedes M. C. Bustamante, Michael T. Coe, Ruth S. DeFries, Michael Keller, Marcos Longo, J. William Munger, Wilfrid Schroeder, Britaldo S. Soares-Filho, Carlos M. Souza & Steven C. Wofsy

Nature **481**, 321–328 (2012)

In the ‘Natural and anthropogenic climatic variation’ section of this Review, we incorrectly referred to the North Atlantic Oscillation as a contributor to the 2005 Amazonian droughts. We should instead have referred to the Atlantic Multidecadal Oscillation, which is an oceanic phenomenon related to anomalies in sea surface temperature in the tropical north Atlantic Ocean and can affect drought in the Amazon Basin. We thank L. Aragao for drawing this error to our attention. This has been corrected in the PDF and HTML versions online.

have formed, or both, in diagenetic processes — those involving iron species made available near the sediment surface as a result of biogeochemical or physical changes that occurred shortly after sedimentation.

More specifically, the authors envisage the formation of nanometre-scale domains consisting of macromolecular complexes formed from organic carbon and iron oxide(hydroxide) minerals, and posit that these domains are sufficiently stable to persist for many thousands of years. Their results indicate that the carbon loading of these complexes may decrease during long-term oxygen exposure, but the overall apparent stability of the complexes, particularly under anoxic conditions, implies that they can provide a mechanism for the long-term sequestration of organic carbon.

Further work is needed to confirm that Lalonde and colleagues' chemical analysis specifically detects only the organic matter bound in iron complexes, given that the method has previously been applied only to studies of soil matrices⁴. But assuming the observations are robust, the authors' findings are impressive from a variety of perspectives. First, they represent a notable advance in our understanding of how organic matter can be preserved in aquatic sediments, one that draws parallels with similar organic-carbon stabilization processes that are prevalent in soils⁵.

Second, the chemical characteristics of the organic matter associated with the iron complexes suggest that it is enriched in biochemical components that are typically considered to be highly degradable. This observation may help to explain the apparent paradox of how intrinsically labile biomolecules can be preserved in the sedimentary record.

Third, if the overall fraction of organic carbon bound in iron complexes is similar to that observed for the samples measured in this study, then such complexes represent a globally important contributor to the long-term sedimentary sink of organic carbon. Taken together, the findings underline the tight coupling between the biogeochemical cycles of carbon and iron.

However, the authors' study¹ leaves many open questions. For example, the precise mechanism by which iron stabilizes organic matter remains unknown, as does the nature and evolution of iron-organic-carbon complexes during early diagenesis and longer-term burial. The selectivity of the stabilization processes for different types of organic compound involved in the formation of iron-organic-carbon complexes, and the impact of this on the legacy of past biological activity preserved as organic signatures in the rock record, also deserve further attention. We can only speculate on how these processes have influenced organic-carbon burial during periods when oceanic conditions were significantly different from today's — for example, because of changes in sea level or concentrations of

dissolved oxygen. Overall, however, the recognition of a 'rusty sink' for organic carbon in ocean sea-floor sediments represents a major advance in our understanding of this enigmatic component of the carbon cycle. ■

Tim I. Eglinton is at the *Geologisches Institut, ETH Zürich, 8092 Zürich, Switzerland.*

AGEING

Sorting out the sirtuins

Debates over the role of sirtuin proteins in ageing are maturing into functional assessments of the individual proteins. It seems that overexpression of a specific sirtuin can extend lifespan in male mice. [SEE LETTER P.218](#)

DAVID B. LOMBARD & RICHARD A. MILLER

Abraham Lincoln once said that God must have loved the common people because he made so many of them. Nature must feel the same way about the sirtuins, a large family of proteins that achieved celebrity status when one member was found to increase lifespan in yeast¹. But are the mammalian sirtuins the rock stars of an ensemble of anti-ageing proteins, or merely members of the entourage? The original model, proposed in about 2005, that sirtuins have broadly evolutionarily conserved roles in promoting longevity per se is now being refined through more detailed functional investigations of each sirtuin². On page 218 of this issue, Kanfi *et al.*³ follow this trend by reporting that overexpression of a sirtuin called SIRT6 leads to a modest extension of lifespan in male, but not female, mice*.

Does an extension of lifespan imply an effect on ageing? Not necessarily: interventions unrelated to ageing, such as giving insulin to a person with type I diabetes, can increase mean and maximal lifespan. The lifespan extension observed by Kanfi and colleagues in SIRT6-overexpressing male mice could be explained, at least partially, by SIRT6 acting as a tumour suppressor. Because male mice have a higher incidence of spontaneous cancer than female mice (incidences of 81% and 50%, respectively, were observed in this study), an anticancer protein (perhaps SIRT6?) would have a larger effect on lifespan in males than in females.

Proving that a lifespan-increasing intervention indeed acts by delaying ageing processes is not a simple matter. For example, acceptance of the idea that lifespan extension by caloric restriction (a diet with reduced calorie intake) reflects a genuine deceleration of ageing emerged gradually from evidence⁴ that restriction slows age-related changes in the

e-mail: timothy.eglinton@erdw.ethz.ch

1. Lalonde, K., Mucci, A., Ouellet, A. & Gélinas, Y. *Nature* **483**, 198–200 (2012).
2. Hedges, J. I. & Keil, R. G. *Mar. Chem.* **49**, 81–115 (1995).
3. Schmidt, M. W. I. *et al. Nature* **478**, 49–56 (2011).
4. Mehra, O. P. & Jackson, M. L. *Clays Clay Miner.* **7**, 317–327 (1958).
5. Kaiser, K. & Guggenberger, G. *Org. Geochem.* **31**, 711–725 (2000).

properties of proliferative and non-proliferative cells in many tissues, and does so in multiple organ systems. Similar cases are being constructed by researchers proposing that dwarf mice could act as models for slowed ageing⁵.

Reports of lifespan increases in mutant or drug-treated mice, particularly studies in which the observed effects are modest, often prove difficult for other laboratories to repeat. This is presumably due to subtle but crucial variations in the animals' diet or genetic background, or in husbandry practices⁶. Moreover, the preference for publication of positive over negative findings inevitably inserts a smattering of false positive results into the literature, and these can be identified only by attempts to replicate experiments. One strength of Kanfi and colleagues' paper³ is that SIRT6 overexpression increased male lifespan in each of two groups of mice, which were derived from two different founder animals. However, the test for maximal lifespan — usually taken as stronger evidence than an effect on median longevity alone — reached statistical significance in only one of the two mouse groups. If the longevity effect seen by the authors proves robust, determining whether SIRT6 overexpression does indeed slow ageing will still require follow-up studies analysing a wide range of age-sensitive endpoints.

In their article, Kanfi and colleagues include some observations hinting at potential mechanisms by which SIRT6 overexpression might affect the lifespan of male mice. Compared with their normal counterparts, SIRT6-overexpressing males had modestly reduced serum levels of the hormone IGF-1, and the signalling activity of IGF-1 receptors was weaker in perigonadal fat tissue in males but not in females. Previous reports have found that SIRT6 attenuates intracellular signalling initiated by IGF-1 and insulin⁷. Furthermore, dramatic deficits in IGF-1 and/or growth hormone (GH, which stimulates IGF-1 secretion)

*This article and the paper³ under discussion were published online on 22 February 2012.

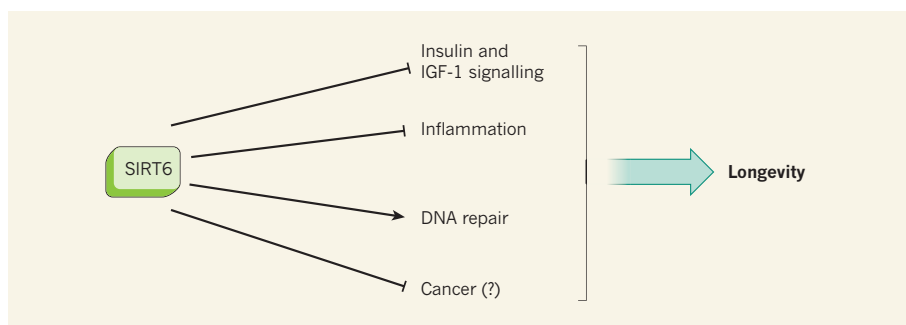


Figure 1 | Potential mechanisms of action of SIRT6 on longevity. Several reports^{7,10–12} have demonstrated effects of the sirtuin protein SIRT6 on the activities of the hormones insulin and IGF-1, as well as on inflammation and DNA repair. These effects, together with a possible delay in cancer progression, could contribute to the increased lifespan in SIRT6-overexpressing male mice reported by Kanfi and colleagues³.

lead to slower ageing and increased lifespan in at least four varieties of mutant mouse⁶. And mutations in the gene encoding the GH receptor in humans are associated with strong protection against diabetes and cancer⁸. So, it is plausible that SIRT6 overexpression in mice might work through blunting of the GH/IGF-1 pathway. Evidence⁹ that rat longevity can be augmented by surgical removal of intra-abdominal — but not subcutaneous — fat has begun to focus attention on metabolic and hormonal effects on specific fat depots as potential levers for pharmacological control of ageing.

It is noteworthy that the effects of SIRT6 overexpression reported by Kanfi *et al.*³ are seen only in male mice. Previous results⁶, by contrast, indicate that mutations in components of the GH/IGF-1 pathway usually have greater effects on longevity in female mice. This apparent discrepancy might be explained by differences between the mice in terms of underlying disease proclivities, levels of sex-specific hormones, inter-animal conflict or fat-tissue biology, leading to gender-specific responses to mutations, drugs and nutritional interventions. Working out the basis for these sex-specific interactions should provide clues to the mechanisms involved in these anti-ageing manipulations, and perhaps even help to answer the vexing question of why women tend to live longer than men.

SIRT6 has other roles that could foster longer lifespan (Fig. 1). It promotes chromosomal stability by several mechanisms, and above-normal SIRT6 expression increases the efficiency of DNA repair¹⁰. SIRT6 also reduces the expression of genes regulated by the NF- κ B and HIF-1 α proteins, which have roles in inflammation, cancer and, potentially, longevity^{11,12}. It will be of interest to assess these aspects of SIRT6's function in mice overexpressing the protein, and to test more definitively whether they contribute to protection against cancer and promotion of longevity.

The recent spate of activity in sirtuin research, now supplemented by the present work, supports the case

for placing the sirtuins on the front line of ageing research, sitting cheek by jowl with other promising contestants, such as the proteins TOR, FoxO, AMPK, NRF2 and ATF4. To paraphrase Winston Churchill, the discoveries of Kanfi *et al.* do not by any means represent the end of sirtuin research, nor even the beginning

of the end. But they are, perhaps, the end of the beginning. ■

David B. Lombard and Richard A. Miller are in the Department of Pathology and the Geriatrics Center, University of Michigan School of Medicine, Ann Arbor, Michigan 48100-2200, USA.
e-mails: davidlom@umich.edu;
millerr@umich.edu

1. Kaeberlein, M., McVey, M. & Guarente, L. *Genes Dev.* **13**, 2570–2580 (1999).
2. Finkel, T., Deng, C.-X. & Mostoslavsky, R. *Nature* **460**, 587–590 (2009).
3. Kanfi, Y. *et al.* *Nature* **483**, 218–221 (2012).
4. Weindruch, R. & Sohal, R. S. *N. Engl. J. Med.* **337**, 986–994 (1997).
5. Bartke, A. *Phil. Trans. R. Soc. Lond. B* **366**, 28–34 (2011).
6. Ladiges, W. *et al.* *Ageing Cell* **8**, 346–352 (2009).
7. Xiao, C. *et al.* *J. Biol. Chem.* **285**, 36776–36784 (2010).
8. Guevara-Aguirre, J. *et al.* *Sci. Transl. Med.* **3**, 70ra13 (2011).
9. Muzumdar, R. *et al.* *Ageing Cell* **7**, 438–440 (2008).
10. Mao, Z. *et al.* *Science* **332**, 1443–1446 (2011).
11. Kawahara, T. L. A. *et al.* *Cell* **136**, 62–74 (2009).
12. Zhong, L. *et al.* *Cell* **140**, 280–293 (2010).

MATERIALS SCIENCE

Continuity through dispersity

By making polymers whose central blocks have a range of lengths, materials have been prepared that contain separate, intermeshed domains extending throughout the material — a highly desirable structure.

RICHARD A. REGISTER

Ideas about nanometre-scale self-assembly — the phenomenon in which nanoscale objects organize themselves into arrays — are commonly underpinned by two assumptions. The first is that the higher the degree of regularity the array has, the better. The second is that obtaining a high degree of regularity requires the assembling objects to be as monodisperse (identical in shape and size) as possible. But these assumptions have recently come under intense scrutiny in the case of self-assembling block copolymers. The molecules in these materials consist of long runs (blocks) made from different monomer types, which can separate into distinct nanoscale domains. Writing in the *Journal of the American Chemical Society*, Mahanthappa and colleagues¹ report that molecules of 'ABA' triblock copolymers (in which A and B represent different blocks) robustly generate desirable self-assembled structures containing separate A and B domains when the middle blocks have a broad range of lengths.

For many applications, it is preferable for

the A and B domains of a block copolymer to be bicontinuous — that is, to interweave and extend throughout the material. For example, in photovoltaic films used in solar cells, excitons (electron–hole pairs, where holes are quasiparticles formed by the absence of electrons) generated by photon absorption could rapidly diffuse to interfaces between bicontinuous A and B domains, thus allowing the charges to separate, find their way out of the film and so generate a current^{2,3}. Or consider the separator membranes in batteries, which divide the anode from the cathode but provide a medium through which ions can pass. A bicontinuous material could provide both a soft, rubbery domain that would facilitate ion diffusion, and a hard, glassy domain that would maintain the membrane's mechanical integrity and stiffness⁴.

Unfortunately, the regular bicontinuous structures (such as the gyroid phase; Fig. 1a) that form in conventional, near-monodisperse copolymers consisting of only two types of block occupy frustratingly narrow slices of 'composition space' — if the fraction of material occupied by A blocks is expressed

► **NATURE.COM**
For more on
sirtuins, see:
go.nature.com/8shahko

BREAD

A taste of history.

BY ELIZA BLAIR

The phone rang, of course, as always, as forever, when her hands were full.

Marli tongued it on from her suit collar and kept kneading, answering belatedly with a distracted “Yaoh?”

It was Kert. “Marli. What is hap?” he said. “No vid?”

“Hands,” Marli said. She was trying to tell if her project looked enough like the tutorial video to finally stop squishing and go wash everything from elbows out. “New hobby. Hold.” She drifted the few feet across her hab and flipped the manual cover on the camera feed with her knee, still forearms-deep in the ball of goo, working it gently with her fingers. A careful toe-push activated the two-way video and sent her floating slowly back towards the kitchen.

“Ewwwww,” said Kert, leaning back from his camera in exaggerated horror. “What is?”

“Bread,” said Marli, enjoying his discomfort.

“Noway. Bread not goo,” Kert objected. “Is flat. Is crunchy. Is from Hub.”

“No Hub monopoly on bread. Plus Hub bread also goo, before.”

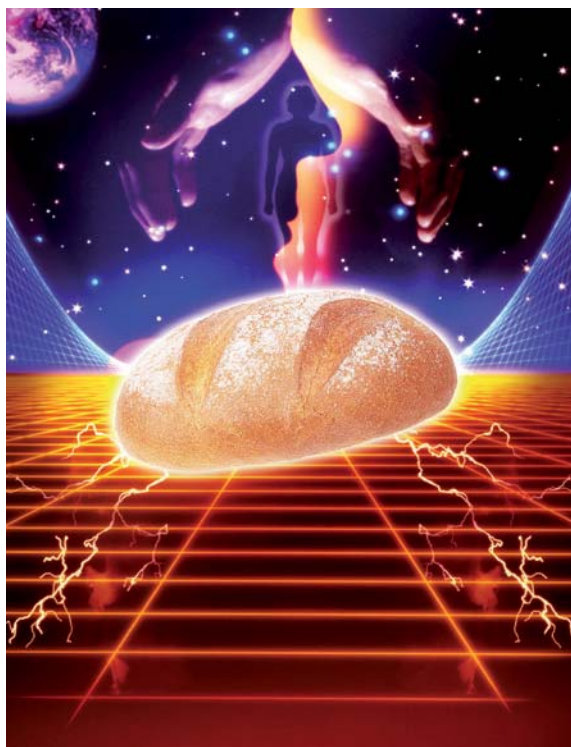
“Before?” Kert raised an eyebrow. “Before cook? But why not ‘fuge, why hands?”

“Fuge not hot. Well, not right hot.” Marli decided her tired fingers had had enough and began carefully extracting herself from the blob. “This bread old kind. Yeast make bubbles. In grav, bubbles rise. No grav, I help.” If she had done it gently enough, her hands had broken the bubbles into smaller bubbles and redistributed them throughout the dough, the second such spreading she had done with this ball since she ‘fuged the ingredients together. The reasons for this were not fully explained in the tutorial, but she thought she understood. She wanted fluffy old-style bread, not a crust of Hub’s hardtack around a centre of air. Words like ‘gluten’ had not yet entered her vocabulary.

“Now what?” Kert asked, as she pushed the ball into the sterilizer and put her hands in the ultrasound.

“Now hold. Tutorial says half hour.”

“Boring,” Kert winked out. Marli laughed. They had been friends for two years, and neighbours for one, and knew each other too well. Kert was intrigued but unwilling to admit it; he would pretend he was too



busy to chat until he showed up at her hatch unannounced. She swiped a little condensation from the bulkhead next to the fridge, where the air circulation was poor, and did her twice-daily readout checklist; but there was nothing much to tidy in the little hab. She read the news on her spex while she waited the requisite half hour. At the appointed time she reached into the sterilizer again, separating the ball into two neat halves, one of which would go in the fridge for a future experiment. The remaining one she formed into a neat ball again, now about the size of the ones at Hub’s gym, and spent some time fiddling with the sterilizer controls.

By the time Kert arrived, his helmet folding neatly into his suit collar even as the airlock finished cycling, she could feel just a hint of heat through the sterilizer’s thick composite window. The dough ball floated in the exact centre, held in place by automatic jets of air.

They hugged, and she gave him a squeezer of weak tea, taking a second from the pantry and thumbing the insta-heat packet. They talked of everything and nothing —

the pirates who had hacked a Hub near L5 and plundered it of supplies, nearly killing the family of four who had

emergency-docked just days later to an empty larder; the upcoming biannual gathering of habbers, when their modular units would come together and connect in huge chains and webs covering a hundred cubic miles of space around L2, using their combined mass to slingshot and nudge each other into new orbits in the beautiful, complicated, computer-guided dance called the Tumble. Where they might go next.

The sterilizer dinged. They were so deep in conversation that Marli had forgotten to check it. She peered at the irregular brown object, larger now, and dubiously opened the hatch as the last of the hot air was sucked away and stored.

“Looks like shit,” Kert observed.

Marli shot him a look as she nudged the crusty ball into the centre of the hab. “Careful, hot. Hold.”

“Again? Hub bread just open packet. Your bread take long, Marli Stewart.”

“Old method. Tradition.” Marli shot him a look as her contextual software belatedly found the reference, then pushed him playfully so they both bounced against the inner hull. “We need remember. Where we come from.”

Kert nudged the ball into the breeze from the air cyclor. Marli could feel the heat in the air that came off it, feel the faint condensation forming again by the fridge, the uncontrolled moisture put off by this anachronism of a food. Watching Kert reposition the ball and finally hold it in place with his fingertips, she felt wild, untamed, connected to the ancient past. This was how her ancestors must have felt, pulling their own creations from stone ovens heated by open flame.

“Stay with me,” she said, unthinkingly, then smiled as her brain caught up. “This cycle. Meet hatches. Double hab it.”

Kert raised an eyebrow at her and drifted forward, holding the bread in his gloved hands. “Is cool now,” he said softly, offering it to her. “Ready?”

Marli nodded and grabbed it around the middle, pushing her fingertips through the thin crust towards the still-scalding centre. Kert smiled and shifted his hands, holding half of the spheroid. Marli gripped the other half, and they braced against each other. “On three.” ■

Eliza Blair has previously appeared in A Science Fiction Omnibus. She can be found online at <http://elizablair.net/elizaeffect>.

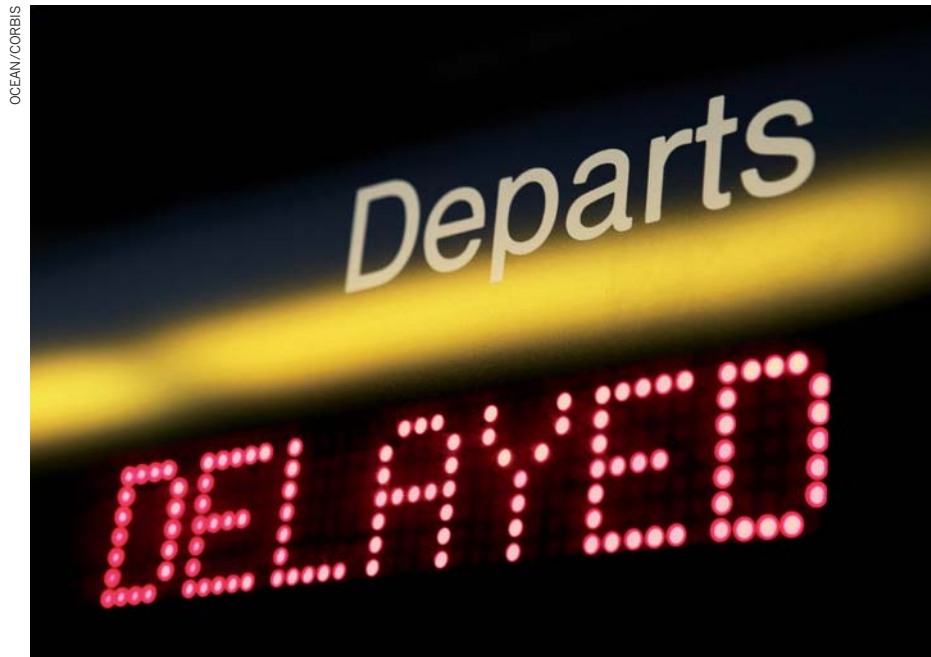
➔ **NATURE.COM**
Follow Futures on
Facebook at:
go.nature.com/mtoodm

CAREERS

COLUMN Postdoc as both a job and career pathway **p.235**

ADVICE FORUM Get expert tips on scientific career issues go.nature.com/nabnle

NATUREJOBS For the latest career listings and advice www.naturejobs.com



RETIREMENT

Sticking around

Academics who delay retirement could create roadblocks for early-career researchers.

BY VIRGINIA GEWIN

Graeme Hugo, a demographer at the University of Adelaide, Australia, doesn't mince his words. He says Australia has inadvertently created a 'lost generation' of academics: those under 40 years old. Hugo says that the academic workforce is now dominated by baby boomers who were hired in droves during university expansions in the late 1960s. More than half of the academic workforce is over 50 years old, so he expects 40% of academics and lecturers to retire over the next decade.

But this is not the ideal situation for young academics it seems to be. Faculty members are delaying their retirement and, increasingly, vacated permanent posts are divided into contractual, non-tenure-track jobs. Tired of navigating the stream of temporary posts, nearly 40% of Australian academics under the

age of 30 plan to leave academia or the country, according to a survey conducted last September (see go.nature.com/rnhl4w).

"The old academic system — the romantic notion of sitting at your desk pondering the nature of the Universe — is gone," says one of the survey's authors Emmaline Bexley, a social scientist at the University of Melbourne. Her results bolster a growing concern that the best and brightest will, increasingly, avoid the insecure, underpaying jobs that seem to dominate academia.

Australia's academic workforce woes are acute, but they also illustrate how chronic global trends are making it increasingly difficult for young academics to establish independent research careers. "Postponed retirement and a general increase in life expectancy, coupled with the continuing shift towards more non-tenure-track and part-time positions, are producing a discouraging

situation for entry-level faculty," says John Curtis, director of research and public policy at the American Association of University Professors (AAUP) in Washington DC.

The huge turnover once expected from baby-boomer retirements doesn't look likely to happen. "Even when academics retire, universities do not necessarily replace tenure-track lines in a one:one ratio," says Paula Stephan, who studies the economics of science at Georgia State University in Atlanta.

The proportion of teaching staff with full-time tenure-track jobs in the United States has dropped by more than half, from 16.1% in 1975 to 7.6% in 2009, according to AAUP data (see 'Employment status of US university teaching staff'). Conversely, the number of full-time non-tenure-track faculty members rose by just under half, from 10.3% to 15.1%. The proportion of part-time faculty members hit a peak of 41.1% in 2009.

Early-career researchers need to be aware that demographics will shape their prospects of a permanent post in academia. For some, non-tenure-track posts, despite concerns about poor job security, offer a viable option and a welcome level of freedom. Others may want to re-evaluate their academic ambitions.

ARRESTED RETIREMENT

Baby boomers will eventually retire. But the predicted retirements may come in a series of swells rather than the surge expected with mandatory retirement, which ended in the United States in 1994. "The pipeline isn't emptying like administrations thought it would," says Cathy Trower, director of the Collaborative on Academic Careers in Higher Education at the Harvard University in Cambridge, Massachusetts. Some academics predict that there will be little turnover of tenured positions in the next two decades. "There is not going to be any massive wave of retirements for a while — we've got 20 years to wait for the anaconda bulge of baby boomers to work itself out," says Frances Rosenbluth, deputy provost for social sciences and faculty development at Yale University in New Haven, Connecticut.

There are exceptions. Administrators at Cornell University in Ithaca, New York, for example, have taken the unusual step of launching a campaign to raise US\$100 million to recruit new faculty members before the expected onslaught of retirements. US federal government agencies, many of which also have an ageing workforce, are preparing for future recruitment needs. For example, ►

► the US Office of Naval Research in Arlington, Virginia, anticipating the retirements, continues to develop its young-investigator programme, this year up to \$8.5 million to attract researchers.

The recession has only exacerbated a growing trend towards deferring retirements; no one expects a retirement boom, even if the economy improves. The Higher Education Generational Survey, conducted by Versta Research for Fidelity Investments in Boston, Massachusetts, a

financial-services corporation that provides retirement savings plans for people employed in the higher-education sector, found that more than half of faculty respondents expect to delay retirement or to never retire. Similarly, a survey of faculty aged 60 years and older by retirement-planning company Teachers Insurance Annuity Association and College Retirement Equities Fund in New York, published in December 2011, found that 60% of respondents would like to and expect to work past the conventional retirement age of 65. And 15% of the respondents say that they will have to continue working past the age of 65 because of financial concerns. "It's a real concern if this 15%, the ones who want to retire but don't feel they can financially, starts to increase — because that has the potential to become a workforce-management issue," says Valerie Conley, a higher-education researcher at Ohio University in Athens.

In the United Kingdom, the latest figures from the Higher Education Statistics Agency in Cheltenham indicate that the number of faculty members aged over 65 at higher-education institutions with science faculty members increased from 970 in 2006 to 1,375 in 2010, a jump of almost 42%. That number is likely to increase, especially if proposed government plans to cut public-sector pension benefits go through. "Changes in pensions could make retiring at a point you might otherwise want less accessible — and in turn that means posts are not as vacant as they might otherwise have been," says Sally Hunt, general secretary of the University and College Union in London.

WORKFORCE REPERCUSSIONS

The University of California, Berkeley, exemplifies what happens when an institution with an excellent benefits package repeals its mandatory retirement age: there is no incentive to leave. The number of older faculty members has grown exponentially since 1994; out of

about 1,500 faculty members, 75 are older than 70 years, according to Marc Goulden, director of data initiatives at the university. As the number of older faculty members has risen, the number of young tenure-track faculty members has fallen: in 1980, roughly 16% of the faculty members were under 35 years old; just 6% were in this age bracket in 2011.

The economic downturn has made finding ways to encourage retirement a hot topic on US campuses where budgets are tight. "The question taking shape seems to be, 'what do we have to do to get them to go?'; but that's the wrong focus," says Goulden. "We can't get them to go now that mandatory retirement no longer exists; we have to build a fluid, flexible system where older faculty can contribute to the university, perhaps even electing to turn over their tenure-track line for another position that is mutually beneficial to all."

Phased-retirement schemes — such as a reduction in the workload over a three-year period in exchange for a commitment to leave — are having the most success in encouraging retirement. Such schemes were in place at 32% of US campuses in 2006, according to an AAUP survey by Conley conducted in 2007 on the changes in faculty retirement policies. Such schemes seem to ease faculty members into the transition towards retirement, and help administrators to manage the campus workforce, says Claire van Ummersen, a senior adviser for the American Council on Education, an organization in Washington DC that represents more than 1,600 campuses.

But with increasing student enrolment, universities are relying more on contract staff to teach. "The increase in universities' use of contingent faculty members to teach is a sign that the trinity of teaching, research and service is breaking down," says Martin Finkelstein, from the department of education leadership, management and policy at Seton Hall University

in South Orange, New Jersey. He says that one emerging reconfiguration of faculty functions is a move towards limited-contract faculty members who teach or conduct research.

Taking a contract position at a 'dream' university, just to get a foot in the door, may be tempting, but may not be the best decision for someone who is eager and able to hold out for a tenure-track post. "Most people who start out in contract appointments stay in contract appointments," says Finkelstein.

For young academics, the most frustrating thing is the lack of job security that comes with fixed contracts. "The huge numbers of staff on short-term contracts remains one of the great secrets of UK higher education, and we must do more to support staff who find themselves continuing to operate without decent job security even after years of training and work within higher education," says Hunt.

TENURE-FREE WILL

Efforts are afoot to improve job security. In 2008, campuses in the United Kingdom, for example, started implementing the Concordat to Support the Career Development of Researchers, coordinated by Britain's research councils, to establish standards for research staff on fixed-term contracts. "There is a move towards universities providing open contracts for research staff, moving away from fixed-term contracts," says Janet Metcalfe, director of Vitae, an organization in London supporting doctoral researcher career development. The number of fixed-term contracts in the United Kingdom has declined in recent years — from 43.2% of positions in 2004 to 33.3% in 2010, according to data from the Higher Education Statistics Agency (see go.nature.com/o9xg2c). Some institutions, including University College London and the University of Bristol, have publicly stated their intentions to move away from fixed-term contracts.



"Universities do not necessarily replace tenure-track lines in a one:one ratio."

Paula Stephen

EMPLOYMENT STATUS OF US UNIVERSITY TEACHING STAFF

Trends in instructional staff employment status in the United States (all institutions, national totals).



There are also moves to create new types of non-tenure-track positions — particularly at medical schools. Van Ummersen says that medical schools are filled with full-time non-tenure-track positions; mainly clinicians who teach and conduct research. Typically, staff are hired through the hospital, rather than by the medical school or university, to keep pace with the hospital workload. The positions come with promotional opportunities, but unlike tenure-track positions, they carry no penalty if the requirements of promotion are not achieved. The University of California's outgoing provost, Lawrence Pitts, expects that universities will follow suit and create different types of faculty positions to meet the needs of the institution. But, given their traditional conservative approach to faculty governance, those changes will be evolutionary rather than revolutionary, he says.

Trower says that not having to worry about tenure can give researchers certain freedoms. "Junior faculty members often don't take as many risks if they need publications to get tenure," she says. She suggests that breakthrough science often comes from places such as interdisciplinary centres at universities, which, unlike departments, are typically filled with non-tenure track positions.

"Young scholars tell me that they are happy on non-tenure-track lines because they have a life and a sense of balance," says Trower. Goulden's data on why early-career researchers elect to leave academia bolsters this idea (M. Mason, D. Goulden and K. Frasch *Academe* 95, 11–17; 2009). "Our survey data show that young researchers believe the hours are long and the possibility of having work–life balance diminishes if one seeks tenure," he says.

Hugo is convinced that, at least in Australia, the prospects for young academics are still good. But, he says, universities need to build more flexibility into recruitment schemes to deal with the constraints imposed by people working longer.

Van Ummersen says that there are opportunities for universities to be creative — for example, establishing 'half-time' tenure-track lines that could be split between early-career researchers and faculty members entering phased retirement agreements — to help meet career and personal issues for both junior and senior faculty members.

"Universities have only begun to change the way they adjust to the needs of faculty members," says van Ummersen. "The university was invented for monks in monasteries. That model doesn't fit the twenty-first century." ■

Virginia Gewin is a freelance writer based in Portland, Oregon.

COLUMN

The postdoc dilemma

Balancing a career and the obligations of a full-time job can be deceptively difficult, says **Gaston Small**.

As a graduate student, I overheard a faculty member advise a colleague to include funding in a grant proposal for a graduate student rather than a postdoctoral fellow. The reason? "Postdocs spend all their time writing papers from their dissertation work," she said. I vowed that, when I became a postdoc, I would not let paper and grant writing disrupt my research in the laboratory.

Fast-forward five years. To my dismay, I do spend much of my time writing and revising manuscripts, many of them from my previous work. I see this as part of the postdoc's primary challenge: balancing a job and a career.

Although I have published six papers from my dissertation work, I have several other manuscripts from the same project competing for space on my desk and weighing on my conscience. Each paper represents collaborations with other early-career scientists, and I am acutely aware that delays affect their careers, as well as my own. I give priority to work related to my current position, but when there are lulls, I dedicate several days to revising one of these old manuscripts in the queue.

Once, at a workshop for early-career researchers, a panellist warned us that it was unethical for postdocs to spend time working on projects other than the one that pays their salary. Yet, when I confessed this transgression to my supervisor, he told me not to worry. I will be writing papers from our current project for years to come, he said. And, hence, will be siphoning time from future projects. It is the perpetual circle of grant-writing life.

The job–career balance is a fundamental challenge for postdocs. Fulfilling the obligations of the project that currently pays your salary is, of course, essential, but at the same time postdocs need to push previous work through the publication process, which often entails multiple revisions. Writing grant applications, and applying and interviewing for faculty jobs are necessary activities; two years of postdoc funding runs out quickly. These additional responsibilities to our careers are as time-consuming as obligations to our full-time jobs.

Perhaps this juggling act is necessary preparation for a career in academia. When I first started as a graduate student, I had a romanticized notion of academics leisurely discussing the fundamentals of ecology with colleagues over a coffee or a beer, sketching graphs on the backs of napkins and filling a blackboard



IMAGEZOO/CORBIS

with equations. The reality has been quite different. The faculty members with whom I've worked are chronically busy. A friendly "how's it going?" to a colleague in the hall is always met with, "oh, really busy this week!". I too have adopted this response. There is always a deadline looming for the next proposal. There are committee meetings, classes to prepare for and students to advise. Responding to e-mails can take the better part of a working day. This routine does not leave much time for pondering big ideas, or catching up on the backlog of old papers that need to be written.

Embracing this culture of academic life is an important part of the postdoc journey. There is often a tension between current job and peripheral obligations. At least part of the solution is clear communication with your supervisor. When applying for a postdoc, you should find out exactly how you are expected to spend your time and check regularly with your supervisor to make sure that your productivity is meeting these expectations.

After two years as a postdoc, I am still learning how to balance these competing obligations. Last week I received an e-mail from a PhD student who worked with me as an undergraduate during my PhD research. He wanted to know the status of a manuscript I am supposed to be revising. I was mortified when I realized that I had not opened the file in fourteen months. Sorry, Pedro. I'll get right on that. ■

Gaston Small is a postdoc in ecology at the University of Minnesota in St Paul.

There are also moves to create new types of non-tenure-track positions — particularly at medical schools. Van Ummersen says that medical schools are filled with full-time non-tenure-track positions; mainly clinicians who teach and conduct research. Typically, staff are hired through the hospital, rather than by the medical school or university, to keep pace with the hospital workload. The positions come with promotional opportunities, but unlike tenure-track positions, they carry no penalty if the requirements of promotion are not achieved. The University of California's outgoing provost, Lawrence Pitts, expects that universities will follow suit and create different types of faculty positions to meet the needs of the institution. But, given their traditional conservative approach to faculty governance, those changes will be evolutionary rather than revolutionary, he says.

Trower says that not having to worry about tenure can give researchers certain freedoms. "Junior faculty members often don't take as many risks if they need publications to get tenure," she says. She suggests that breakthrough science often comes from places such as interdisciplinary centres at universities, which, unlike departments, are typically filled with non-tenure track positions.

"Young scholars tell me that they are happy on non-tenure-track lines because they have a life and a sense of balance," says Trower. Goulden's data on why early-career researchers elect to leave academia bolsters this idea (M. Mason, D. Goulden and K. Frasch *Academe* 95, 11–17; 2009). "Our survey data show that young researchers believe the hours are long and the possibility of having work–life balance diminishes if one seeks tenure," he says.

Hugo is convinced that, at least in Australia, the prospects for young academics are still good. But, he says, universities need to build more flexibility into recruitment schemes to deal with the constraints imposed by people working longer.

Van Ummersen says that there are opportunities for universities to be creative — for example, establishing 'half-time' tenure-track lines that could be split between early-career researchers and faculty members entering phased retirement agreements — to help meet career and personal issues for both junior and senior faculty members.

"Universities have only begun to change the way they adjust to the needs of faculty members," says van Ummersen. "The university was invented for monks in monasteries. That model doesn't fit the twenty-first century." ■

Virginia Gewin is a freelance writer based in Portland, Oregon.

COLUMN

The postdoc dilemma

Balancing a career and the obligations of a full-time job can be deceptively difficult, says **Gaston Small**.

As a graduate student, I overheard a faculty member advise a colleague to include funding in a grant proposal for a graduate student rather than a postdoctoral fellow. The reason? "Postdocs spend all their time writing papers from their dissertation work," she said. I vowed that, when I became a postdoc, I would not let paper and grant writing disrupt my research in the laboratory.

Fast-forward five years. To my dismay, I do spend much of my time writing and revising manuscripts, many of them from my previous work. I see this as part of the postdoc's primary challenge: balancing a job and a career.

Although I have published six papers from my dissertation work, I have several other manuscripts from the same project competing for space on my desk and weighing on my conscience. Each paper represents collaborations with other early-career scientists, and I am acutely aware that delays affect their careers, as well as my own. I give priority to work related to my current position, but when there are lulls, I dedicate several days to revising one of these old manuscripts in the queue.

Once, at a workshop for early-career researchers, a panellist warned us that it was unethical for postdocs to spend time working on projects other than the one that pays their salary. Yet, when I confessed this transgression to my supervisor, he told me not to worry. I will be writing papers from our current project for years to come, he said. And, hence, will be siphoning time from future projects. It is the perpetual circle of grant-writing life.

The job–career balance is a fundamental challenge for postdocs. Fulfilling the obligations of the project that currently pays your salary is, of course, essential, but at the same time postdocs need to push previous work through the publication process, which often entails multiple revisions. Writing grant applications, and applying and interviewing for faculty jobs are necessary activities; two years of postdoc funding runs out quickly. These additional responsibilities to our careers are as time-consuming as obligations to our full-time jobs.

Perhaps this juggling act is necessary preparation for a career in academia. When I first started as a graduate student, I had a romanticized notion of academics leisurely discussing the fundamentals of ecology with colleagues over a coffee or a beer, sketching graphs on the backs of napkins and filling a blackboard



IMAGEZOO/CORBIS

with equations. The reality has been quite different. The faculty members with whom I've worked are chronically busy. A friendly "how's it going?" to a colleague in the hall is always met with, "oh, really busy this week!". I too have adopted this response. There is always a deadline looming for the next proposal. There are committee meetings, classes to prepare for and students to advise. Responding to e-mails can take the better part of a working day. This routine does not leave much time for pondering big ideas, or catching up on the backlog of old papers that need to be written.

Embracing this culture of academic life is an important part of the postdoc journey. There is often a tension between current job and peripheral obligations. At least part of the solution is clear communication with your supervisor. When applying for a postdoc, you should find out exactly how you are expected to spend your time and check regularly with your supervisor to make sure that your productivity is meeting these expectations.

After two years as a postdoc, I am still learning how to balance these competing obligations. Last week I received an e-mail from a PhD student who worked with me as an undergraduate during my PhD research. He wanted to know the status of a manuscript I am supposed to be revising. I was mortified when I realized that I had not opened the file in fourteen months. Sorry, Pedro. I'll get right on that. ■

Gaston Small is a postdoc in ecology at the University of Minnesota in St Paul.